ABSTRACT

Title of Dissertation: STRUCTURAL STUDY OF THE EARLY SECRETORY PROTEIN B AND ATYPICAL RIO2 KINASE

Soheila Bahmanjah, Doctor of Philosophy, 2015

Dissertation Directed By: Associate Professor Nicole LaRonde
Department of Chemistry and Biochemistry

Mycobacterium tuberculosis, Mtb, is a successful pathogen that secretes variety of proteins to manipulate the host defense mechanisms and create a proper environment for its survival. The progression of the disease depends on the regulated secretion of essential virulence factors such as culture filtrate protein-10 (CFP-10), and early secretory antigenic target-6 (ESAT-6). Mtb regulates the secretion of the virulence factors through interactions of MycP1 protease with EspB (Early secretory protein B) protein.

One of the goals of this thesis was to provide the first crystal structure of EspB protein in order to gain insight into the expression, secretion, function and transmembrane translocation of this protein. EspB292 (residue 1-292) structure has 16 monomers in the asymmetric unit that are arranged into a set of four homotetramers. To clarify the components that affect oligomerization of the EspB292 in the biologically relevant conditions, we set up experiments that mimic the phagosomal environment.

In addition to work done on EspB protein, structural determination of atypical *Chaetomium thermophilum* Ct-Rio2 in presence of transition state analogs such as sodium orthovanadate and beryllium fluoride is reported to define the importance of individual domains in binding ATP. Rio2 is required for site D cleavage and maturation of the small subunit ribosomal RNA (rRNA). In this study, the structures of the Ct-Rio2 and its binding to ADP/BeF_2 and ADP/VO_4^{2-} are investigated. The structure are expected to mimic the transition state of the phosphoryl transfer from ATP to Asp257 in Rio2's active site and the subsequent hydrolysis of the aspartyl phosphate that could power late cytoplasmic 40S subunit biogenesis.

Collaborative work on the structural characterization of the G228 Ct-Rio2 mutant protein is also reported. Our goal is to investigate the evolutionary relationship between Rio kinases and the canonical eukaryotic protein kinases. Our hypothesis is that canonical eukaryotic protein kinases have evolved from Rio kinases. We have made several mutations in a key differentiating position of the kinase domain sequence of Ct-Rio2, G228, and have collected biochemical data to test the ability of the mutated Rio kinase to autophosphorylate and carry out ATP hydrolysis.

STRUCTURAL STUDY OF THE EARLY SECRETORY PROTEIN B AND ATYPICAL RIO2 KINASE

By

Soheila Bahmanjah

Dissertation Submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in Partial Fulfillment
of the Requirements for the degree of
Doctor of Philosophy
2015

Advisory Committee:
Professor Nicole LaRonde, Chair
Professor Steve Rokita
Professor Daniel Falvey
Professor Herman O. Sintim
Professor Roy Mariuzza, Dean's Representative

Preface

During my PhD program, I have worked in two labs. During my time in Prof. Jeffery T. Davis lab, I started the study of the amphiphilic natural products for the purpose of anion transport across the lipid bilayer of the membrane. For information on this project the reader is referred to the following research article.

Bahmanjah, S., Zhang, N., Davis, J. T. Monoacylglycerols as transmembrane Cl^- anion transporters. *Chem. Commun.* 2012, *48*, 4432-4434.

In Dr. Nicole LaRonde's lab, I worked on three main projects that are described in detail in the next seven chapters.

Dedication

First and foremost, I would like to dedicate this work to the memory of my father and my brother. They were both immensely proud that I was working toward a PhD. I am sorry that they did not get to see my completed dissertation and I did not get to see them during my PhD program. It is with pride and affection that I dedicate this dissertation to them.

I would also like to dedicate this work to my mother. She never went to school but is immensely talented and well-mannered. She who is not able to read or write did her best so that all her children would be able to do both. She supported us to get the best possible education and I'm forever thankful for that.

Acknowledgements

I would like to express my deepest appreciation to my advisor, Dr. Nicole LaRonde. I was very fortunate to join her research group in May of 2012. She continually guided me in this journey and was very patient, supportive and motivating. I believe Nicole is one of the most inspiring female role models I've ever met in my life. She is very friendly, intelligent, determined, talented, honest, reliable, generous, patient and ambitious. She personally, trained me for most of the experiments I performed in the lab and provided insights to solve multiple research issues. Without her guidance and persistent help this dissertation would not have been possible.

I would also like to thank Dr. Tinoush Moulaei for sharing his knowledge with me and for providing guidance and help. I'm heavily indebted to him for so many encouraging and insightful discussions we had in the lab and I'm forever thankful for his support.

I would also like to take this opportunity and thank all members of the lab, past and present who provided help and made it fun working in the lab.

I would like to thank all the members of this thesis committee for their continued guidance support and especially for devoting so much time to get together in our second candidacy meeting. Dr. Rokita taught me a lot through his lectures on catalysis and through our meetings over the past years. He has certainly been the best teacher I've ever had in my life. I am grateful to Dr. Falvey and Dr.Sintim for their support and understanding. They were so kind and supportive to stay as members of my candidacy committee although I had to change majors from organic chemistry to biochemistry. Dr. Falvey devoted a lot of time to me, asking insightful questions during our meetings and has been very helpful and considerate in our meetings. Many thanks to Dr. sintim for being very responsive and helpful. I believe he

is a genius when it comes to organic chemistry and I've been very fortunate to be one of his students.

Finally, I am grateful to my entire old and new families: my parents, my brothers, and my sisters who stood by me and supported me in this effort. They encouraged me to continue to be successful and work hard to reach this goal. I have to thank my brother, Jaber for his support and love.

And Last, but certainly not the least, I have to thank Mehdi and his family. They have been very supportive of me and have always been by my side when I needed them. I have to thank you Mehdi for putting up with me throughout these years. You mean the world to me.

Table of Contents

Abstract **I**

Preface **V**

Dedication **VI**

Acknowledgements **VII**

Table of Contents **IX**

List of Tables **XIII**

List of Figures **XVI**

Abbreviations **XXV**

Chapter 1: Introduction **01**

1.1 Tuberculosis Causes, Symptoms and Diagnosis **02**

1.2 Tuberculosis Virulence Mechanism **03**

1.3 Mtb Virulence Factors **04**

1.4 Mtb MycP1 and EspB **05**

1.5 Early Secretory Protein B **05**

1.6 Objectives of Structure Determination of EspB **06**

1.7 Ribosome Structure and Function **07**

1.8 Ribosome Biogenesis **08**

1.9 The Involvement of Rio1 and Rio2 in 20S Pre-rRNA Modification **09**

1.10 Rio2 as an ATP-Dependent rRNA Binding Protein **10**

1.11 Objectives **11**

Chapter 2: Purification and Crystallization of EspB **22**

2.1 Expression of EspB292 **23**

2.1.1 Expression of EspB in the Presence of Selenium **23**

2.1.2 Expression of EspB in the Absence of Selenium **24**

2.2 Purification of EspB292 and FL-EspB **24**
2.3 Crystallization of EspB292 and FL-EspB **26**
2.4 Data Collection and Structure Determination **27**
2.5 Overview of the Structure of EspB **28**
2.6 EspB can Adopt a Homodimer Fold **28**
2.7 EspB can Target the EccCb1 ATPase **29**
2.8 Secretion of EspB/Rv3879c Follows Pairwise Dependency **29**
2.9 Inner Membrane Translocation of EspB/Rv3879c Pair **30**
2.10 Secretion Regulation of CFP-10/ESAT-6 **31**
2.11 Summary **32**

Chapter 3: Oligomeric State Characterization of EspB **46**

3.1 Type VII Secretion System **46**
3.2 Role of EspB in Type VII Secretion System **46**
3.3 EspB292 Oligomerization in Literature **47**
3.4 Results and Discussion **48**
3.4.1 EspB292 Oligomerization **48**
3.4.2 BS3 Cross-Linking Protocol **48**
3.4.3 EspB Oligomerization at High Concentration **49**
3.4.4 EspB292 Assembly Formation in Presence of Heavy Metal Ion **49**
3.5 Iron and TB **51**
3.6 Zinc, Copper and TB **51**
3.7 EspB Oligomerization in the Presence of Iron and Copper **52**
3.8 Summary **52**

Chapter 4: Purification and Crystallization of Ct-Rio2 **65**

4.1 Expression of the Ct-Rio2 Protein **65**
4.2 Purification of Rio2 from *Chaetomium Thermophilum* **66**
4.3 Ct-Rio2 Binds to RNA Nonspecifically **67**

4.4 Crystallization of Ct-Rio2 in Presence of ADP **68**
4.5 Crystallization of Ct-Rio2 in Presence of ADP/$BeCl_2$/KF **69**
4.5.1 Data Collection and Structure Determination **70**
4.6 Crystallization of Ct-Rio2 in Presence of ADP/Na_2VO_4 **71**
4.6.1 Data Collection and Structure Determination **71**
4.6.2 Crystallization of Ct-Rio2 in Presence of ADP/Na_2VO_4 at pH 7 **72**
4.6.3 Data Collection and Structure Determination **72**
4.7 Summary **73**

Chapter 5: Structural Analysis of AMP- and ADP-Bound Ct-Rio2 **93**

5.1 Introduction **93**
5.2 Obtaining the Crystal Structure of Ct-Rio2 **93**
5.3 Crystal Structure of Ct-Rio2 with ADP, $BeCl_2$/KF **94**
5.4 The structural Analysis of the AMP-Bound Ct-Rio2 **94**
5.5 Crystal Structure of Ct-Rio2 in the Presence of ADP and Na_2VO_4 **95**
5.6 Summary **96**

Chapter 6: Evolution of the Eukaryotic Protein Kinases **104**

6.1 Evolutionary Origins of the Eukaryotic Protein Kinases **104**
6.2 The Rio Family **104**
6.3 The Common Classification of Proetin Kinases **105**
6.4 Purification of Rio2 Mutants from Chaetomium Thermophilum **106**
6.5 Crystallization of Ct-Rio2 Mutants **106**
6.6 Crystal Structure of AMP-Bound G228N **107**
6.7 Crystal Structure of AMP-Bound G228A **107**
6.8 Crystal Structure of AMP-Bound G228D **108**
6.9 Results and Discussion **108**
6.10 ATPase Activity and Autophosphorylation Comparison **108**
6.11 Summary **109**

Chapter 7: Conclusion and Future Work **122**

7.1 Crystal Structure of EspB **122**

7.2 Crystal Structure of Ct-Rio2 with AMP and ADP/Na_2VO_4 **126**

7.3 The Evolutionary Origins of the Eukaryotic protein Kinases **127**

Bibliography **129**

List of Tables

Table 2.1 The Se-Met EspB292 Scalelog file. The crystal diffracted to 2.5Å. **38**

Table 2.2 Se-Met EspB292 data collection and refinement statistics. **39**

Table 4.1 Data collection and current refinement statistics for AMP bound Ct-Rio2 at the pH of 5.4. **78**

Table 4.2 AMP-bound Ct-Rio2 Scalepack Logfile at pH 5.4. The Ct-Rio2 crystals were formed in the presence of ADP/$BeCl_2$/KF in the hanging drop. The crystals were produced in 0.1M HEPES pH 7.0. 30% v/v jeffamine ED-2001 pH 7.0 at the 2:1 protein to well solution ratio. **79**

Table 4.3 AMP-bound Ct-Rio2 phenix.xtriage logfile at pH 5.4. The Ct-Rio2 crystals were formed in the presence of ADP/$BeCl_2$/KF in the hanging drop. The crystals were produced in 0.1M HEPES pH 7.0. 30% v/v jeffamine ED-2001 pH 7.0 at the 2:1 protein to well solution ratio. **80**

Table 4.4 Data collection and current refinement statistics for AMP bound Ct-Rio2 at the pH of 7. The Ct-Rio2 crystals were formed in the presence of ADP/$BeCl_2$/KF in the hanging drop. The crystals were produced in 0.1M HEPES pH 7.0. 30% v/v jeffamine ED-2001 pH 7.0 at the 2:1 protein to well solution ratio. **82**

Table 4.5 AMP-bound Ct-Rio2 Scalepack Logfile at pH 7. The Ct-Rio2 crystals were formed in the presence of ADP/$BeCl_2$/KF in the hanging drop. The crystals were produced in 0.1M HEPES pH 7.0. 30% v/v jeffamine ED-2001 pH 7.0 at the 2:1 protein to well solution ratio. **83**

Table 4.6 AMP-bound Ct-Rio2 phenix.xtriage logfile at pH 7. The Ct-Rio2 crystals were **84**

formed in the presence of ADP/$BeCl_2$/KF in the hanging drop. The crystals were produced in 0.1M HEPES pH 7.0. 30% v/v jeffamine ED-2001 pH 7.0 at the 2:1 protein to well solution ratio.

Table 4.7 Data collection and current refinement statistics for AMP-bound Ct-Rio2 at the pH of 5.4: The Ct-Rio2 crystals formed in the presence of ADP/Na_2VO_4 in the hanging drop at pH 5.4: The crystals were produced with 0.05M ammonium acetate, 0.01 magnesium chloride, 0.05M TRIS hydrochloride pH 7.5, 10% v/v 2-methyl-2,4-pentandiol **87**

Table 4.8 AMP-bound Ct-Rio2 phenix.xtriage logfile at the pH of 5.4: The Ct-Rio2 crystals formed in the presence of ADP/Na_2VO_4 in the hanging drop at pH 5.4: The crystals were produced with 0.05M ammonium acetate, 0.01 magnesium chloride, 0.05M TRIS hydrochloride pH 7.5, 10% v/v 2-methyl-2,4-pentandiol. **88**

Table 4.9 AMP-bound Ct-Rio2 Scalepack Logfile at pH 7. The Ct-Rio2 crystals formed in the presence of ADP/Na_2VO_4 in the hanging drop at pH 5.4: The crystals were produced in 0.05M ammonium acetate, 0.01 magnesium chloride, 0.05M TRIS hydrochloride pH 7.5, 10% v/v 2-methyl-2,4-pentandiol. **91**

Figure 4.10 Crystal structure of Ct-Rio2 bound to ADP and transition state analog sodium orthovanadate: The crystal structure of Ct-Rio2 shows a bilobal kinase domain (green and yellow) connected to an N-terminal winged helix domain (pink). The “hinge” region connects the N- and C-terminal lobes of the kinase domain. The ADP binding loop “P-loop”, metal binding loop used to bind Mg^{2+}, and catalytic loop which contains the catalytic D232 are indicated by arrows. The cartoon representation was made by PYMOL. **92**

Table 6.1 Data collection and current refinement statistics for AMP-bound G228N. The Ct-Rio2 crystals formed in the presence of ATP in the hanging drop at pH 5.4. **114**

Table 6.2 Data collection and current refinement statistics for AMP-bound G228A: The Ct-Rio2 crystals formed in the presence of ATP in the hanging drop at pH 5.4. **115**

List of Figures

Figure 1.1 *Mycobacterium tuberculosis* and skin test: TB is an infectious disease caused by Mtb. According to the skin test results one third of the world population are infected by TB. **13**

Figure 1.2 Global tuberculosis report: In 2014, 8.7 million chronic cases of TB were reported, leading to1.4 million deaths. Geographically, the burden of TB is highest in Asia and Africa. India and China together account for 40% of the world's TB and Africa has 24% of the world's TB infection. **14**

Figure 1.3 TB pathophysiology: 1) Infection via inhalation of droplets; 2) Ingestion of organisms by macrophages; 3) Multiplication within macrophages; 4) Minimal inflammatory response; 5) Extension of organisms into the bloodstream; 6) Hypersensitivity reaction; 7) Development of clinical infection. **15**

Figure 1.4 The Mtb's cell wall is consisted of several layers including phthiocerol dimycocerosate (DIM/DIP) layer, mycolic acids, arabinogalactan (AG), and peptidoglycan, lipoarabinomannan, and phosphatidylinositol mannoside. **16**

Figure 1.5 T7SS is encoded by ESX-1: ESX-1 is a specialized protein secretion system that is encoded by genes of region of difference 1 (RD1) and its extended regions. RD1 is comprised of nine genes (Rv3871 to Rv3879c), which are deleted from the primary cause of attenuation of the M*ycobacterium bovis* bacille Calmette Guérin (BCG) vaccine strain. That is attributed to a deletion of nine genes from the ESX-1 locus. **17**

Figure 1.6 The representation of rRNA maturation: The 90S complex is cleaved on the 35S pre-rRNA at processing sites A_0, A_1 and A_2, within the large 90S complex, to produce the pre-40S and the pre-60S particles. The pre-40s particle is rapidly exported to the cytoplasm but the pre-60S particle remains in the nucleolus and undergoes a complex **18**

maturation pathway in the nucleoplasm and is subsequently, delivered to the cytoplasm.

Figure 1.7 Formation of ribosomal beak: In the cytoplasm, the major structural rearrangement that occurs on the pre-40S is the formation of the ribosomal "beak" structure that requires a cascade of phosphorylation and dephosphorylation events. Cryo-EM and biochemical studies have revealed that the formation of the beak is concomitant with the release of assembly factors such as Ltv1 and Enp1. **19**

Figure 1.8 Schematic representation of 40S subunit maturation in *Saccharomyces cerevisiae*: Pre-40S particles at different stages of assembly are purified using TAP to obtain the systematic information about cytoplasmic synthetic pathway of maturation. In the cytoplasm, the 20S pre-rRNA is first, methylated at the 3´ end of the 18S rRNA sequence by the function of Dim1. **20**

Figure 1.9 Schematic representations of ribosomal protein and rRNA binding sites: In the 40S structure model, the Rio2-binding site is located near the Rpss15. Leger-Silvestre et al have reported that the pre-40S ribosome that lacks Rps15 fail to bind Rio2 and be exported to cytoplasm. In bacteria, the Rps18 and Rps16 homologues (S13 and S9) are phosphorylated at serine and threonine residue. **21**

Figure 2.1 Schematic representation of EspB constructs: EspB constructs are designed with an N-terminal 6x His tag followed by a TEV cleavage site. **34**

Figure 2.2 Preparative gels of EspB292 after first His-trap column (A) and size exclusion chromatography (B): Lane M is the protein ladder. Gel of EspB292 displays pure protein at the expected molecular weight of 31.8 kDa. **35**

Figure 2.3 Preparative gels of FL-EspB after size exclusion chromatography: Lane M is the protein ladder. Gel of EspB292 displays pure protein at the expected molecular weight of 37.8kDa. **36**

Figure 2.4 EspB292 crystals: Diamond-shaped crystals at pH 8.5 that diffracted to 2.5Å. **37**

Figure 2.5 Overall structure of EspB292: EspB292 has 16 monomers in the asymmetric unit that are arranged into a set of four homotetramers. **40**

Figure 2.6 Structure of EspB has four homotetramers: Two sets of these tetramers (tetramers 1 and 2; tetramers 3 and 4) are arranged parallel to each other. **41**

Figure 2.7 Structure of EspB has four homotetramers: Contact is observed between the C-terminal loop of one tetramer and the C-lobe helices of the adjacent tetramer. **42**

Figure 2.8 Overall monomeric structure of EspB: The monomer is mostly helical and cylindrical in overall shape. **43**

Figure 2.9 Interface between monomers A and D: Black dashed lines are hydrogen bonds and red spheres are waters. Interface between the C-lobe helices (α4, α5 and α6). **44**

Figure 2.10 Interface between monomers A and B: Black dashed lines are hydrogen bonds and red spheres are waters. Interface between the C-lobe helices (α4, α5 and α6). **45**

Figure 3.1 Proposed model of T7SS that is encoded by ESX-1: The phagosomal signals initialize the transcription and secretion of two pairs EspA/EspC and CFP-10/ESAT-6. Once expressed, Rv3868 assembles these pairs into multimers that are subsequently recognized by Rv3871 and Rv3870. These two FtsK family proteins form a translocase pore that the multimeric protein assembly is likely to pass through. Thereafter, the interaction of the N-terminal gate-like domain of Rv3877 with MycP1, a serine protease, opens the pore, thus enabling translocation of the substrate assembly from cytoplasm to the periplasm-like space. Ultimately, the substrate assembly crosses the mycomembrane through a channel formed by Rv3869, Rv3882c and the four proteins of MCE cluster 1 (Mce1B, Mce1C, Mce1F and Rv0177). **55**

Figure 3.2 EspB292 SDS gel in presence of cross-linker: BS3 with respective spacer arm length of 11.4Å was used as the cross-linking agent that could covalently link two interacting proteins. Lanes 3 to 13 display the formation of protein assemblies as we decreased the crosslinker concentration (from 1.92mM to 53µM). To all lanes except lane 1 BME was added as the reducing agent. The protein concentration was kept constant at 53µM in all lanes. **56**

Figure 3.3 EspB292 SDS gel in presence of cross-linker: Lanes 1 to 9 displays the formation of protein assemblies as we increased the protein concentration (1µM-100µM). To all lanes except lane 10 BME was added as the reducing agent. Lanes 10 and 11 are showing the pure protein without the crosslinker. Numbers between panels indicate sizes in kilodaltons. The protein:Crosslinker ratio was kept constant at in all lanes. **57**

Figure 3.4 Mass spectrometry of EspB292 proteins: The MS displays the formation of higher order complex assemblies at 50µM. **58**

Figure 3.5 Phagosomal acidification: The acidification of phagosome is established by the action of the vacuolar H^+-ATPase (v-ATPase) and the complementary roles played by other transporters and channels to counterbalance the ionic gradient of H^+ across the membrane. For example the influx of anions such as Cl^- via either the CFTR or the ClC family channels is the major counter ion transport that maintains vacuolar acidification **59**

Figure 3.6 Iron, zinc, and copper concentration: Early phagosome contains high concentration of Fe^{3+} which decreases within the next 24 hours after phagocytosis. The decrease in iron III ion concentration happens along with the increase in the concentrations of Cu^{2+}, and Zn^{2+} ions as well as Sulphur and chlorine. Additionally, salts, metals and trace elements can impact the antimicrobial mechanisms and the biology of the pathogens. **60**

Figure 3.7 EspB292 assembles to higher order aggregation complexes upon increasing the concentration of copper from 1mM to 120mM: The native gel of 20µM EspB at all lanes in presence of increasing concentrations of $CuCl_2$. **61**

Figure 3.8 Denaturing gel of 20μM EspB and 2.5mM BS3 cross-linker at all lanes in presence of increasing concentrations of $CuCl_2$. 62

Figure 3.9 EspB292 assembles to higher order aggregation complexes upon increasing the concentration of iron (III) from 10μM to 10mM: The native gel of 20μM EspB at all lanes in presence of increasing concentrations of $FeCl_3$. 63

Figure 3.10 Denaturing gel of 20μM EspB and 2.5mM BS3 cross-linker at all lanes in presence of increasing concentrations of $FeCl_3$. 64

Figure 4.1 Schematic representation of Ct-Rio2 construct: Ct-Rio2 construct is designed with an N-terminal 6x His tag followed by a TEV cleavage site. 74

Figure 4.2 Purification of Ct-Rio2 on His Trap™ Column: Supernatant containing overexpressed Ct-Rio2 was passed over one His Trap™ column. The SDS-PAGE gels, stained with Coomassie Blue, is labeled with lysate (Lys), supernatant (Sup), and flowthrough (FT) the fraction numbers 1-7 that represent 30 mM imidazole concentration and 9-13 represent 50mM imidazole at which the fractions eluted. Ct-Rio2 begins to elute from the His TrapT^M column at 30mM imidazole. 75

Figure 4.3 Preparative gels after size exclusion chromatography: Lane M is the protein ladder. Gel of Ct-Rio2 displays pure protein at the expected molecular weight of 44.6 KDa. 76

Figure 4.4 The Ct-Rio2 crystals formed in the presence of ADP/$BeCl_2$/KF in the hanging drop at pH 5.4: The crystals were produced in 0.1M HEPES pH 7.0., and 30% v/v jeffamine ED-2001 pH 7.0 at the 2:1 protein to well solution ratio. 77

Figure 4.5 The CtRio2 crystals formed in the presence of ADP/$BeCl_2$/KF in the hanging drop at pH 7: The crystals were produced in 0.1M HEPES pH 7.0. 30% v/v jeffamine ED-2001 pH 7.0 at the 2:1 protein to well solution ratio. 81

Figure 4.6 Alignment between bound AMP to CtRio2 at both pH 5.2 and pH 7.0: The structures of Ct-Rio2 are aligned based on the AMP location. The orientation of Asp 257 is the same for both of the crystals. **85**

Figure 4.7 The Ct-Rio2 crystals formed in the presence of ADP/Na_2VO_4 in the hanging drop at pH 5.4: The crystals were produced with 0.05M ammonium acetate, 0.01 magnesium chloride, 0.05M TRIS hydrochloride pH 7.5, 10% v/v 2-methyl-2,4-pentandiol. **86**

Figure 4.8 Crystal Structure of AMP-bound Ct-Rio2: The crystal structure of Ct-Rio2 shows a bilobal kinase domain (green and orange) connected to a N-terminal winged helix domain (pink). The "hinge" region connects the N- and C-terminal lobes of the kinase domain. **89**

Figure 4.9 The Ct-Rio2 crystals formed in the presence of ADP/Na_2VO_4 in the hanging drop at pH 7.0: The crystals were produced in 0.05M ammonium acetate, 0.01 magnesium chloride, 0.05M TRIS hydrochloride pH 7.5, 10% v/v 2-methyl-2,4-pentandiol. **90**

Figure 4.10 Crystal structure of Ct-Rio2 bound to ADP and transition state analog sodium orthovanadate: The crystal structure of Ct-Rio2 shows a bilobal kinase domain (green and yellow) connected to an N-terminal winged helix domain (pink). The "hinge" region connects the N- and C- terminal lobes of the kinase domain. The ADP binding loop "P-loop", metal binding loop used to bind Mg, and catalytic loop which contains the catalytic D232 are indicated by arrows. The cartoon representation was made by PYMOL. **92**

Figure 5.1 Crystal Structure of AMP bound Ct-Rio2 shows bilobal structure: The crystal structure of Ct-Rio2 shows a bilobal kinase domain (green and orange) connected to an N-terminal winged helix domain (pink). The "hinge" region connects the N- and C-terminal lobes of the kinase domain. The ATP binding loop "P-loop", metal binding loop used to bind Mg, and catalytic loop which contains the catalytic D257 are indicated by arrows. The cartoon representation was made by PYMOL. **98**

Figure 5.2 Crystal Structure of AMP bound Ct-Rio2: The N-terminus of Ct-Rio2 contains a winged helix domain. It is comprised of four α helices followed by a β-strand wing β-strand motif, and another α-helix. AMP binds in the large cleft between the two lobes. The N-terminal lobe consists of a five-stranded antiparallel β sheet and a long α helix. The C-terminal lobe of the Rio domain consists of two antiparallel β strands and two α helices. **99**

Figure 5.3 Polar Contacts with the AMP: Crystal structure of AMP bound Ct-Rio2 shows that the α phosphate of AMP makes polar contacts with Asp 257, Asn 234 and Mg ion. **100**

Figure 5.4 Alignment between bound AMP to CtRio2 at both pH 5.2 and pH 7.0: The structures of Ct-Rio2 are aligned based on the AMP location. The orientation of Asp 257 is the same for both of the crystals. **101**

Figure 5.5 The active site of Ct-Rio2 bound to ADP and Orthovanadate: The crystal structure of Ct-Rio2 with vanadate and ADP shows that polar Contacts can be observed between α and β phosphate of ADP with Lys. The β phosphate of ADP is simultaneously bound to vanadate oxygens, $Mg^{2+}$and Asp257. **102**

Figure 5.6 The overlay of the active site of Ct-Rio2 bound to ADP and Orthovanadate vs. Ct-Rio2 bound to AMP: The Asp 257 displays a dramatic movement from the AMP-bound complex to the vanadate bound complex. **103**

Figure 6.1 Testing of the expression of Ct-Rio2 protein and its G228 mutants in E.Coli: The induced (I) samples show a single band at 46 kD which corresponds to Ct-Rio2 and its mutant. The uninduced (U) samples didn't display the same band. **111**

Figure 6.2 Preparative gels after size exclusion chromatography: Lane M is the protein ladder. Gel of Ct-Rio2 displays pure protein at the expected molecular weight of 44.6 KDa for all of the mutants. **112**

Figure 6.3 Ct-Rio2 mutants' crystals: A. G228A B. G228N and C. G228D Grown with **113**

ATP/Mg^{2+}.

Figure 6.4 Crystal Structure of AMP-bound G228N: The crystal structure of Ct-Rio2 with Asn mutation at residues 228 shows a bilobal kinase domain (blue and yellow) connected to a N-terminal winged helix domain (green). The "hinge" region connects the N- and C-terminal lobes of the kinase domain. **116**

Figure 6.5 Crystal Structure of AMP-bound G228A: The crystal structure of Ct-Rio2 with alanine mutation at residues 228 shows a bilobal kinase domain (green and orange) connected to an N-terminal winged helix domain (pink). The "hinge" region connects the N- and C- terminal lobes of the kinase domain. **117**

Figure 6.6 The overlay of the C-Terminal domain in the crystal structures of G228N and G228G (wild type Ct-Rio2): The winged helix domains of the two structures are not completely superimposable. The presence of Asn in the structure pushes the winged helix domain to provide a larger phosphate binding loop. **118**

Figure 6.7 The overlay of the AMP domain in the crystal structures of G228N and G228G (wild type Ct-Rio2): The N and C terminal lobes of the two structures are not completely superimposed. The presence of Asn in the structure pushes the winged helix domain back to provide a larger phosphate binding loop. **119**

Figure 6.8 The overlay of the C-Terminal domain in the crystal structures of G228N, G228A and G228G (wild type Ct-Rio2): The winged helix domain of the G228A and G228 are very close to each other compared to the G228N. The presence of Asn in the structure seems to push the winged helix domain back to provide a larger phosphate binding loop. **120**

Figure 6.9 The proteins with different point mutations at 228, show very different activities in their A) ATPase and B) Auto-phosphorylation experiments: The G228A and Ct-Rio2 with G228 are similarly active in the hydrolysis of ATP but mutant's residues such as those mostly observed in epks (G228V, and G228R) are less active. In line with our **121**

predictions, the auto-phosphorylation of Ct-Rio2 mutants follows an opposite order.

Abbreviations

A	Alanine
Å	Angstrom
ADP	Adenosine-5'-diphosphate
APS	Advanced Photon Source
ATP	Adenosine-5'-triphosphate
BME	β-mercaptoethanol
C	Cysteine
CFP-10	culture filtrate protein-10
COOT	Crystallographic Object Oriented Toolkit
Ct-Rio2	*Chaetomium thermophilum* Rio2
D	Aspartic Acid
Da	Dalton
°C	Degrees Celsius
ClC	Chloride Channel
CGV	Calmette-Guérin vaccine
DNA	Deoxyribonucleic Acid
E	Glutamic Acid
EDTA	Ethylenediaminetetraacetic Acid
ePKs	Eukaryotic Protein Kinases
ESAT-6	early secretory antigenic target-6
EspA	early secretory protein A
EspB	Early secretory protein B
EspC	early secretory protein C
ESX-1	early secretory export type I
F	Phenylalanine
FPLC	Fast Protein Liquid Chromatography
Ftsk	Filamenting temperature-sensitive mutant K
G	Glycine
H	Histidine

HIV	human immunodeficiency virus
HYSS	hybrid substructure search
ITS	Internal Transcribed Spacer
I	Isoleucine
IPTG	Isopropyl-β-D-thiogalactopyranoside
K	Lysine
L	Leucine
M	Molar
M	Methionine
MAD	Multiple Wavelength Anomalous Diffraction
MDR-TB	multiple drug-resistant tuberculosis
Mg	Milligram
$MgCl_2$	Magnesium Chloride
Mg^{2+}	Magnesium ion
mL	Milliliter
mM	Millimolar
Mtb	*Mycobacterium tuberculosis*
MycP1	Mycobacterium Protease 1
N	Asparagine
NECAT	Northeastern Collaborative Access Team
P	Proline
PCR	Polymerase Chain Reaction
PEG	Polyethylene Glycol
PHENIX	Python-based Hierarchical Environment for Integrated Xtallography
Q	Glutamine
R	Arginine
RB	Ribosome Biogenesis
RD1	region of difference 1
RMSD	Root Mean Square Deviation
RNA	Ribonucleic Acid
rRNA	Ribosomal RNA
S	Serine

S	Svedburg Unit
SAD	Single Wavelength Anomalous Diffraction
SeMet	Selenomethionine
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SSU	Small Ribosomal Subunit
T	Threonine
T7SS	Type VII secretion system
TAP	Tandem Affinity Purification
TB	*tubercle bacillus*
TEV	Tobacco Etch Virus
TLR	Toll-like receptors
Tris	Tris(hydroxymethyl)aminomethane
UV	Ultraviolet
V	Valine
W	Tryptophan
wHTH	Winged Helix Domain
Y	Tyrosine
µL	Microliter
µM	Micromolar

Chapter 1: Introduction

In this study, X-ray crystallography is utilized as a form of high resolution microscopy to visualize protein structures at the atomic level. Using the three dimensional structure of a protein, we have obtained insight into the function and mechanism of our target proteins. Armed with this information that is presented in the following chapters, our ultimate goal is to design novel drugs or inhibitors that target a particular enzyme or rationally engineering an enzyme for a specific industrial process. To reach these ultimate goals, the interest of my thesis is the structural determination of two classes of proteins so-called EspB and Rio kinase protein.

The diffraction from a single molecule would be too weak to be measurable. So, we use an ordered three-dimensional array of molecules, in other words a crystal, to magnify the signal. Even a small protein crystal might contain a billion molecules. If the crystal is well ordered, then diffraction will be measurable at high angles and resolution and a detailed structure should result. Hence, to obtain structural information about a particular protein, essentially a pure sample of the target protein is required.

We can do this either by isolating the protein from a source, or by cloning its gene into a high expression system and purifying the protein. Following the later process, EspB and Rio2 family of proteins are purified. The pure and homogeneous samples of the proteins have been further used to be crystalized using the hanging drop crystallization method. Our structural observations have provided us with insight into how our target proteins interact with other molecules or one another, how they undergo conformational changes, and how they perform catalysis in the case of Rio kinases.

The first half of this dissertation, chapters 2 and 3, involves the structure determination of the EspB protein. EspB is involved in the virulence of tuberculosis disease. The investigation of

the crystal structure of EspB has been an important key to understand the function of the early secretory protein B in the Mtb progression.

In an additional work, in chapters 4-6, we were interested in structure determination of Ct-Rio2 family of proteins.

1.1 Tuberculosis Causes, Symptoms and Diagnosis

Tuberculosis[1], or TB (short for *tubercle bacillus*), is a contagious disease that attacks the respiratory system. In the past TB was called phthisis, phthisis pulmonalis, or consumption. This disease is caused by *Mycobacterium tuberculosis* (Mtb) (**Figure 1.1**). *Mycobacterium tuberculosis (Mtb)* was first discovered in 1882 by Robert Koch. Tuberculosis generally attacks the lungs and less commonly affects other parts of the body[2]. Mtb spreads through the air when people who have an active TB infection transmit their respiratory fluids in the air by coughing, or sneezing[3]. Most infections do not have symptoms which are known as latent tuberculosis. However, about one in ten latent infections eventually progresses to active disease which, if left untreated, kills more than 50% of those infected[4]. The classic symptoms of active TB infection are chronic cough with bloody sputum, fever, night sweats, and weight loss.

Diagnosis of active TB relies on chest x-rays and radiology, as well as microscopic examination of body fluids[5]. Diagnosis of latent TB relies on the tuberculin skin test (TST) or blood tests (**Figure 1.1**). In 2014, 8.7 million chronic cases of TB were reported, leading to 1.4 million deaths. Geographically, the burden of TB is highest in Asia and Africa. India and China together define 40% of the world's TB and Africa has 24% of the world's TB infections[4] (**Figure 1.2**). Treatment of TB requires usage of multiple antibiotics over a long period of time. Social contacts are also limited during treatment. One of the main problems of TB treatment is the antibiotic resistance in multiple drug-resistant tuberculosis (MDR-TB) infections. Prevention is

always better than treatment. Therefore, vaccination with the bacillus Calmette-Guérin vaccine (CGV) is often required in early ages to prevent the development of TB[6].

Infection of TB starts via the inhalation of the contaminated droplet. Ingestion of the Mtb organisms by macrophages and its multiplication within macrophages leads to the inflammatory responses. The TB can further develop to a more severe disease by the extension of organisms into the bloodstream followed by the hypersensitivity reaction (**Figure 1.3**). Upon Mtb infection, a phagosome is formed via phagocytosis of the bacterium by macrophages[7]. The phagosome undergoes maturation by first fusing with various endosomes via the endocytic pathway followed by fusion with lysosomes to form phagolysosomes. During phagolysosome formation, degradative enzymes that destroy invading pathogens are released and the pH drops to about 4.5. This pH provides an environment that is not conducive for Mtb survival. This process, phagolysosome formation, is one of the multiple antimicrobial responses that the host cell employs to limit the growth of infection[8].

1.2 Tuberculosis Virulence Mechanism

Remarkable capacity of Mtb to infect, and manipulate variety of host defense environments has turned tuberculosis to the world's leading infectious disease after HIV, causing two million deaths every year[4]. Survival of Mtb in macrophages is critical to the virulence and pathogenesis of the bacterium. One survival response that Mtb utilizes in inhibition of phagosome maturation is prevention of acidification of the phagosome by lysosomes to pH of about 4.5. The phagosome is hence, arrested in the endocytic pathway under a promotive pH of about 6.4 and therefore, phagolysosomes are not formed[9].

The phagosome inhibition mechanism is not well understood but studies do show another survival mechanism which is the involvement of a specialized secretion pathway known as ESX-

1. This secretion pathway aids Mtb to manipulate its environment by delivering virulence factors during infection[10]. The secretion of variety of proteins that evade, modulate and control host immune responses creates a favorable environment for its intracellular survival and replication of Mtb. Since *Mycobacteria* have highly protective, and hydrophobic waxy coating of mycolic acids known as the mycomembrane[11], the export of the virulence factors, either to the extracellular environment or directly in the host cell, requires a special system that is known as the ESX or type VII secretion (T7S) pathway (**Figure 1.4**). Type VII secretion system (T7SS) has first been identified in the pathogenic organism *Mycobacterium tuberculosis* H37Rv and the corresponding gene clusters[11]. The T7SS follows a pairwise dependency of proteins for both secretion and function.

The secreted proteins such as culture filtrate protein-10 (CFP-10), and early secretory antigenic target-6 (ESAT-6) that are essential for the virulence of *Mycobacteria* in T7SS are encoded by genes of RD1 (region of difference 1) and extended regions (**Figure 1.5**). Therefore, the deletion of RD1 region attenuates the virulence of TB significantly[12]. This region contains genes encoding ATP dependent motors, a number of transmembrane proteins, a protease (MycP1) and secretory proteins such as CFP-10, ESAT-6, EspA, EspB, EspC, and EspR that are prominent targets of the immune system in humans infected with Mtb[13]. Based on the analysis of RD1 components, previous studies have proposed the following models for the translocation of proteins (**Figure 1.6**).

The phagosomal signals initialize the transcription and secretion of two pairs EspA/EspC and CFP-10/ESAT-6[14]. Once expressed, Rv3868 assembles these pairs into multimers that are subsequently recognized by Rv3871 and Rv3870[15]. These two FtsK family proteins form a translocase pore that the multimeric protein assembly is likely to pass through. This translocase

pore facilitates the ATP-dependent translocation of the substrates through the inner membrane pore protein (Rv3877)[16]. Thereafter, the interaction of the N-terminal gate-like domain of Rv3877 with MycP1, a serine protease, opens the pore enabling translocation of the substrate assembly from cytoplasm to the periplasm-like space[17]. Ultimately, the substrate assembly crosses the mycomembrane through a channel formed by Rv3869, Rv3882c and the four proteins of MCE cluster 1 (Mce1B, Mce1C, Mce1F and Rv0177).

1.3 Mtb Virulence Factors

Various activities have been ascribed to the ESX-1 substrates ESAT-6 and CFP-10, including inhibition of phagosome maturation and cytokine signaling by infected macrophages, interaction with the macrophage immune receptor TLR and inhibition of TLR signaling and formation of pores in mycobacterial phagosome, perhaps allowing bacterial spread[18]. To perform the attributed activities, ESAT-6 and CFP-10 secretion is finely tuned by the interactions of MycP1 and Rv3881c (Early secretory protein B, EspB).

1.4 Mtb MycP1 and EspB Proteins

MycP1 plays a dual role in regulating the secretion activity of ESX-1[19]. Whereas the MycP1 interaction with Rv3877 opens the pore to enable protein secretion, its protease activity negatively regulates secretion via Rv3881c (EspB) to balance the virulence and maintain the long-term Mtb infection[20].

1.5 Early Secretory Protein B

Rv3881c (Early secretory protein B, EspB) is a conserved alanine- and glycine-rich protein that is necessary for inhibiting phagosome maturation[21]. The translocation of EspB through the membrane involves the same machinery used by the other secretory proteins (CFP-10/ESAT-6 and EspA/EspC). EspB initially, pairs up with Rv3879c and co-secretes with CFP-

10/ESAT-6 through Rv3877 to the periplasm-like space. Subsequently, MycP1 cleaves EspB and disrupts the EspB/Rv3879c pair, resulting in the secretion of EspB. The protein MycP1 cleaves EspB at two sites, Ala358 and Ala386 on the C-terminus[22].

1.6 Objectives of Structure Determination of EspB

The T7SS and the ESX-1 secretion pathway is not fully understood and some unknowns that interest my current research include:

- The molecular structure of EspB
- How do EspB monomers interact with one another?
- How does EspB oligomeric state in the crystal relate to its biological functions?
- How does the structure provide insight into cellular environments that Mtb encounters?
- How does the structure of EspB predict the interaction with virulence factors?
- How does the structural information predict the EspB membrane transport mechanism?

In the chapters 2 and 3, I will describe the work that went towards structural determination of EspB and presents the first crystal structure of EspB along with structural analysis.

In chapter two of this dissertation, the purification and crystallization procedure of EspB is explained in detail. We optimized purification of EspB to obtain highly pure protein samples that were immediately used for crystallization. In chapter two, we provide the first crystal structure of residues 1-292 of (EspB292) protein (which is an elastase stable fragment of full length EspB) and in chapter three the structure of EspB is summarized. Our findings suggest that this protein has 16 monomers in the asymmetric unit that is arranged into a set of four homo-tetramers and therefore, it forms high order tetrameric aggregations upon cleavage at the periplasmic space of the mycobacterium by MycP1. Since the cleavage of EspB is essential for

the regulation of virulence factor's secretion, we studied the phagosomal environmental conditions to investigate the elements that would impact the oligomerization of EspB.

In the second half of this dissertation, **chapters 4-6**, as mentioned earlier, we were interested in structure determination of Ct-Rio2 family of proteins. The structural information helps us understand the mechanism of phosphoryl transfer from ATP to these kinases and obtain evolutionary information about the relationship between the Rio family kinases and eukaryotic protein kinases.

1.7 Ribosome Structure and Function

Rio2 is a serine kinase that is involved in the processing of 20S pre-rRNA during the maturation of ribosome[23]. The actual role of them in the Ribosome biogenesis (RB) is not clear yet but recently, we have obtained many interesting information about their functions[24].

Ribosome biogenesis is the process of making ribosomes. It is an incredibly complex process in the cell that spends a lot of energy to provide cell with molecular machinery for protein production. Ribosome biogenesis is critical for cell cycle progression and cellular proliferation in both normal cells and cancer cells. Any misregulation in the folding and assembly process of involved ribosomal RNA (rRNA) or proteins in this process can lead to several diseases.

Mature ribosomes are composed of two subunits in all organisms, the large subunit (LSU) and the small subunit (SSU)[25]. In eukaryotic cells such as *Saccharomyces cerevisiae* (Sc) the LSU is consisted of three rRNAs known as 25S, 5.8S and 5S[26]. In prokaryotic cells such as *Escherichia coli*, there are only two rRNAs in the LSU naming 23S and 5S[27]. The SSU of all of the ribosomes have only one rRNA which is called 18S in eukaryotic cells and 16S in the prokaryotic cells[28].

In any organism, all of the rRNAs are first synthesized as a single transcript[29]. This rRNA transcript will then get cleaved and processed at different sites to produce the mature rRNAs. During this process many ribosomal proteins assemble on the rRNA transcript to generate the mature ribosome. Besides ribosomal proteins and rRNAs, numerous transacting factors are involved in ribosome to promote the modifications, folding and processing of ribosomal precursors and their ordered assembly[30]. Many of these non-ribosomal factors are members of different families of enzymes including ATP-dependent RNA helicases, AAA-ATPases, GTPases, and kinases[31]. The yeast *Saccharomyces cerevisiae* (Sc) and a thermophilic fungus called *Chaetomium thermophilum* (Ct) are representatives of suitable eukaryotic model organisms to study ribosome assembly.

Over the past 20 years, the easy genetic, biochemical and cell biological experimental accessibility has enabled us to obtain significant information about the biogenesis. For example, the tandem affinity purification method (TAP) facilitates the isolation of numerous non-ribosomal intermediates along the ribosomal maturation pathway, which correspond to snap shots of pre-ribosomal particles[27]. The current challenge is to identify the direct interaction partners of the proteins and obtain structural and functional details of the pre-ribosomal particles[32].

1.8 Ribosome Biogenesis

The transcription of ribosomal genes and production of pre rRNA transcript in eukaryotes takes place in the nucleolus[33]. Initially, the RNA polymerase I generates a precursor rRNA (so-called 35S pre-rRNA) that contains the sequences for the mature 18S, 5.8S, 5S and 25S rRNA (**Figure 1.6**). The 35S pre-rRNA is enclosed by external transcribed spacer (5′-ETS and 3′-ETS) and is divided by the internal transcribed spacers (ITS1, and ITS2). Thereafter, ribosomal

proteins assemble on the 35S pre-rRNA and generate a 90S complex that sediment at 90S on a sucrose gradient[34]. The newly formed 90S complex is then cleaved on the 35S pre-rRNA at processing sites A_0, A_1 and A_2, within the large 90S complex, to produce the pre-40S and the pre-60S particles[29]. While the pre-40s particle is rapidly exported to the cytoplasm, the pre-60S particle that contains the 27SA2 pre-rRNA remains in the nucleolus and undergo a complex maturation pathway in the nucleoplasm and is subsequently, delivered to the cytoplasm. The pre-40S complex contains the 20S pre-rRNA[23].

In the cytoplasm, the major structural rearrangement that occurs on the pre-40S is the formation of the ribosomal "beak" structure that requires a cascade of phosphorylation and dephosphorylation events[35]. Cryo-EM and biochemical studies have revealed that the formation of the beak is concomitant with the release of assembly factors such as Ltv1 and Enp1[36] (**Figure 1.7**).

1.9 The Involvement of Rio1 and Rio2 in 20S Pre-rRNA Modification

There are two assembly factors called Enp1 and Ltv1 that are shown to be required for the export of pre-40S complex[37]. The Rio2 presence and binding in the nucleus is required for the export and release of Enp1 and Ltv1 from pre-40S ribosome in the cytoplasm[38]. In the cytoplasm, the 20S pre-rRNA is first, methylated at the 3′ end of the 18S rRNA sequence by the function of Dim1 (**Figure 1.8**). Subsequently, the methylated rRNA is cleaved by the function of endonuclease Nob1 to generate mature rRNA[39]. For the 3′ end access of Nob1 and the cleavage of 20S pre-rRNA, Dim1 and a GTPase called Tsr1 are required to be released from the pre-rRNA[40].

Based on the UV cross-linking and cDNA analysis results by David Tollervey[41] and coworkers, both Dim1 and Tsr1 bind to sequences in the 18S rRNA that includes the decoding

center and regions required for tRNA association[42]. Thus, Dim1 and Tsr1 prevent the pre-40S complex from accessing tRNA and participating in translation. Moreover, RNA assembly factors have shown to play an important role in regulating RNA folding events that occur during the 40S assembly. For Instance, the ChemModSeq analyses has recently revealed that pre-40S particles that contain Rio2, Tsr1, Ltv1 and Enp1 binding sites have generally more flexible particles and rRNA that the late pre-40S complexes that lacked these proteins[43]. It appears that assembly factors such as Ltv1 and Enp1 act as RNA chaperones that maintain a more flexible conformation in the pre-rRNA by delaying folding steps of the 3′ major domain. At certain points during the maturation pathway, these assembly factors are no longer required and they are released by Hrr25-phosphorylation which reduces their affinity for RNA[44].

Several evidences have indicated that Enp1 and Ltv1 play an important role in recruiting Rps3, Rps15 and Rps20[31]. Shortly after the release of Enp1 and Ltv1 the Rio2 and Tsr1 are also released and Rio1 enters the assembly pathway (**Figure 1.9**). However, the exact function of Rio1 is still unclear. Finally, data from the Tollervey lab has confirmed that the Rio1-associated particle in the late stages of maturation pathway contains Dim2 and Nob1[39, 45].

1.10 Rio2 as an ATP-Dependent rRNA Binding Protein

As mentioned earlier, the Rio2 is required for the release of hEnp1 and hLtv1 from pre-40S ribosome in the cytoplasm[46]. Interestingly, the **physical presence** of Rio2 but **not its kinase activity** is necessary for the release of hEnp1 from cytoplasmic 40S precursor. In fact, Rio2 kinase[47] activity is not essential in ribosome assembly[48]. A Rio2 mutant in which the kinase activity is abolished can still support growth but only when overexpressed. We believe that the persistent binding of Rio2 to the pre-40S complex may allow maintenance of a phosphorylated

state even in the presence of competing phosphatase activity. Alternatively, Rio2 may simply be an ATP-dependent RNA-binding protein, perhaps modulated by its own autophosphorylation[48].

Furthermore, incubation of Rio2 containing pre-40S particles with radiolabeled ATP in yeast resulted in no labeling of the pre-40S subunit, suggesting that there are no substrates for the Rio2 protein on the pre-40S particle[49]. In eukaryotes, it is well established that the functional role of the Rio kinases, especially Rio1 and Rio2, is the maturation of the small subunit of the ribosome[50]. In *Saccharomyces cerevisiae* (budding yeast) both Rio kinases are essential and deletion of either one is lethal.

With the goal of elucidating the function of Rio kinases, for several years scientists have been looking for Rio kinase family substrates. Despite years of study in this area, the identification of protein substrates for these enzymes has been elusive[50]. However, all of the Rio kinases that have been purified and characterized have shown autophosphorylation activity.

A recent crystal structure of a eukaryotic Rio2 kinase from *Chaetomium thermophilum* revealed that ADP and a phosphoaspartate intermediate state was generated after incubation with ATP and magnesium ions[48]. This led to the observation of ATPase activity for this enzyme. The ATPase activity is measured to have turnover rate of 0.9 min^{-1} which is higher than that measured for autophosphorylation activity by 50-100 fold. These data, along with the observation that the Rio2 kinase likely binds the pre-40S particle with its active site inaccessible to protein substrate, led to the conclusion that Rio2 functions as an ATPase in the maturation steps of the pre-40S.

1.11 Objectives

The ultimate goal of chapters 5 and 6 is to understand the mechanism of the phosphoryl transfer from the ATP to the Asp residue of Ct-Rio2 in the catalytic cycle. We obtained the

crystal structure of Ct-Rio2 in presence of sodium orthovanadate that mimics the transition state analog of phosphoryl transfer.

The ultimate goal is to regulate cell proliferation by possibly modulating the rate of ribosome biogenesis. Towards this ultimate goal, the focus of my research was to target the non-ribosomal protein, Ct-Rio2 by asking the following questions:

- What is the structure of Ct-Rio2 in presence of transition state analogs?
- What is the function of catalytic loop in the active site of this protein?
- What is the mechanism of the phosphoryl transfer from the ATP to the Asp residue?
- Can Rio2 act as the evolutionary origins of the eukaryotic protein kinases?
- How does the structure of Ct-Rio2 predict the interaction with the ribosomal particles?
- How does Rio2 protein interact with ATP?
- How does the Rio2 protein undergo conformational changes and perform catalysis?

To address these questions, I solved the crystal structure of Ct-Rio2 with the AMP and ADP bound in presence of sodium orthovanadate[51]. I will describe this complex in chapter 6. In chapter 7, I will focus on functional and mutational studies of Ct-Rio2 and describe whether the Rio kinases can be potential evolutionary origins for the canonical eukaryotic protein kinases.

Figure 1.1 *Mycobacterium tuberculosis* and skin test: TB is an infectious disease caused by *Mycobacterium tuberculosis* (Mtb). According to the skin test results one third of the world population are infected by TB[4].

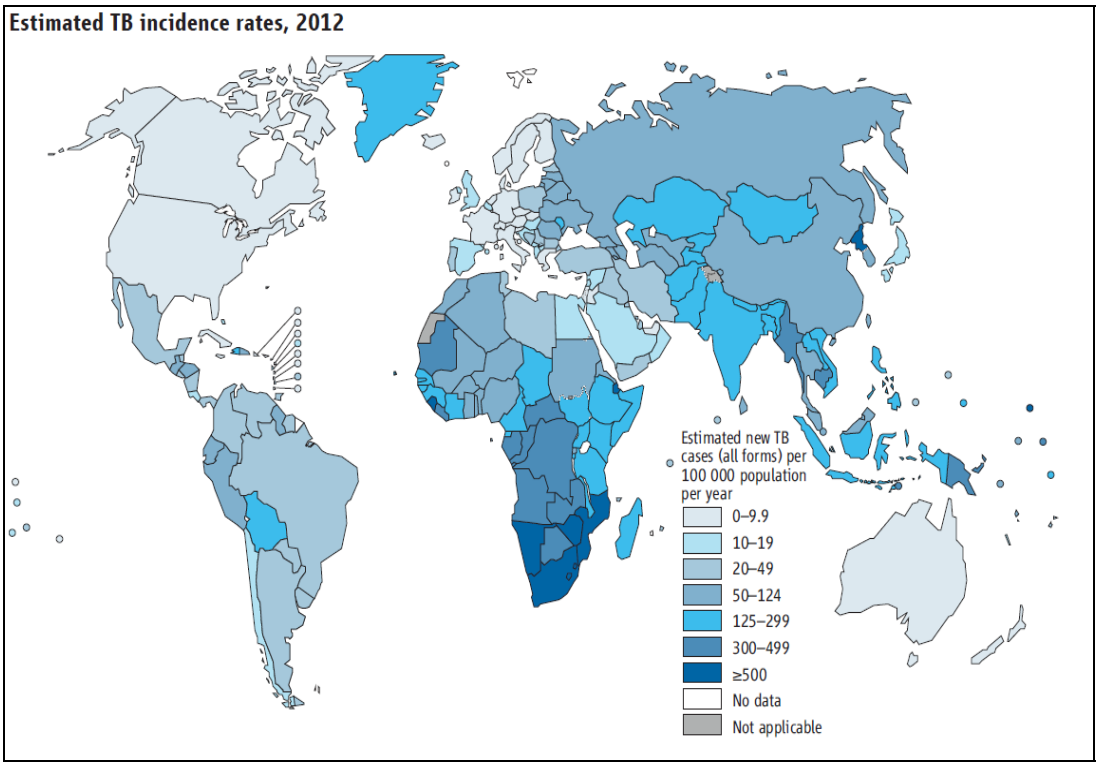

Figure 1.2 Global tuberculosis report: In 2014, 8.7 million chronic cases of TB were reported, leading to1.4 million deaths. Geographically, the burden of TB is highest in Asia and Africa. India and China together account for 40% of the world's TB and Africa has 24% of the world's TB infection[4].

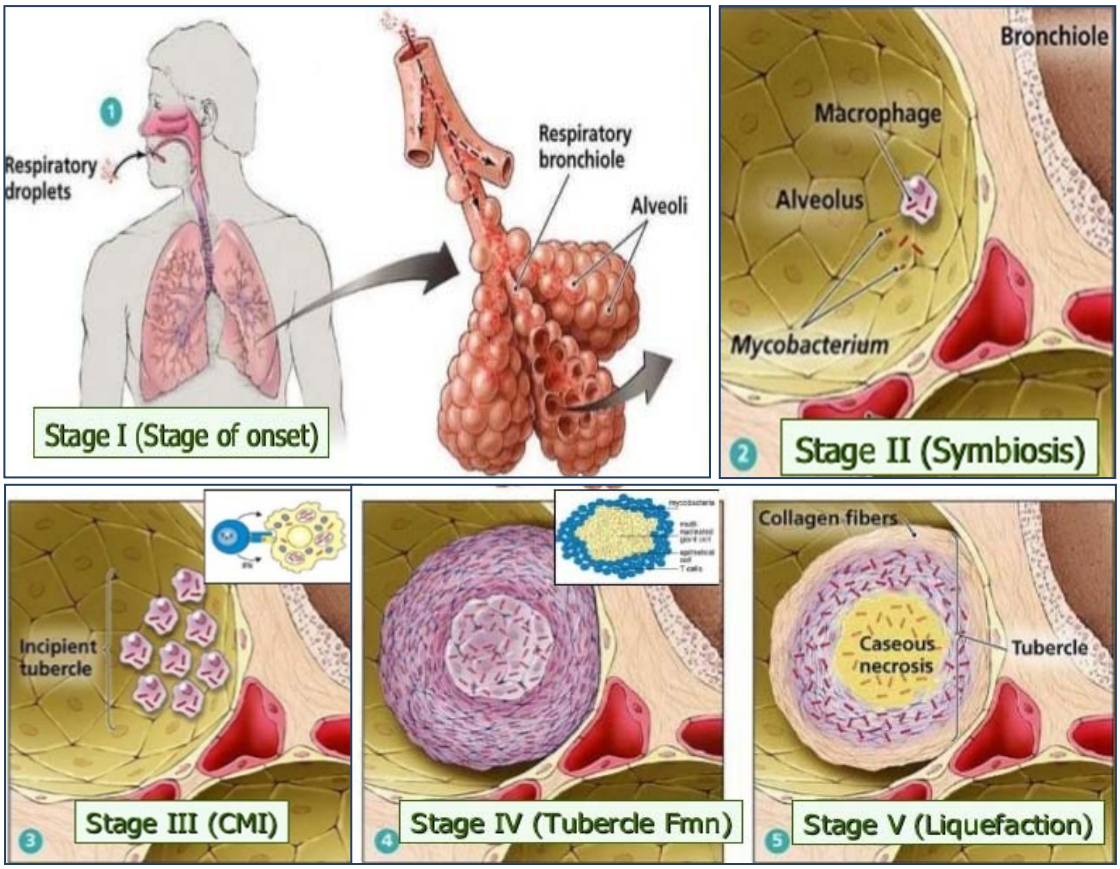

Figure 1.3 TB pathophysiology: 1) Infection via inhalation of droplets; 2) Ingestion of organisms by macrophages; 3) Multiplication within macrophages; 4) Minimal inflammatory response; 5) Extension of organisms into the bloodstream; 6) Hypersensitivity reaction; 7) Development of clinical infection[1].

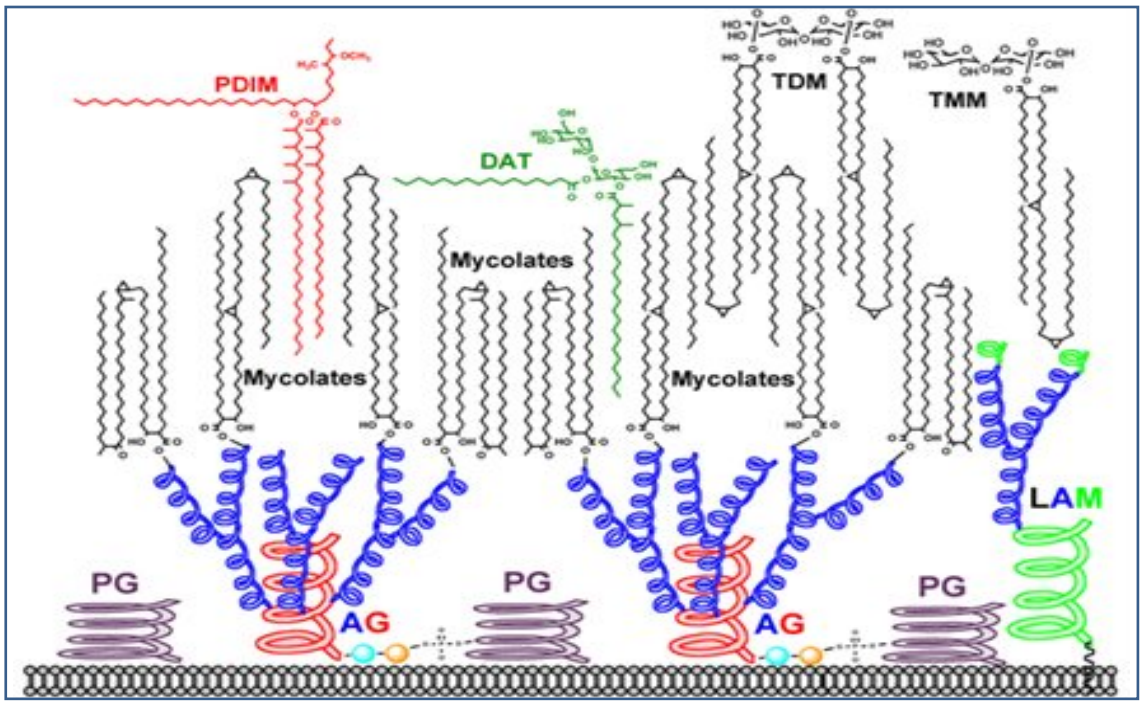

Figure 1.4 The Mtb's cell wall is consisted of several layers including phthiocerol dimycocerosate (DIM/DIP) layer, mycolic acids, arabinogalactan (AG), and peptidoglycan, lipoarabinomannan, and phosphatidylinositol mannoside[11].

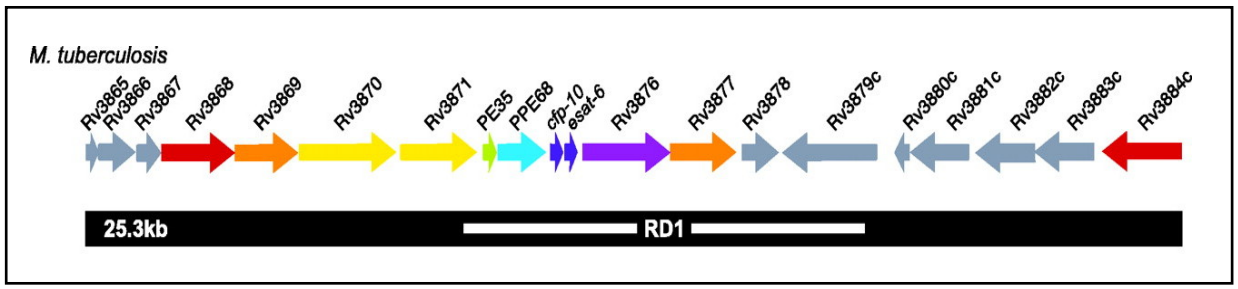

Figure 1.5 T7SS is encoded by ESX-1: ESX-1 is a specialized protein secretion system that is encoded by genes of region of difference 1 (RD1) and its extended regions. RD1 is comprised of nine genes (Rv3871 to Rv3879c), which are deleted from the primary cause of attenuation of the *mycobacterium bovis* bacille Calmette Guérin (BCG) vaccine strain. That is attributed to a deletion of nine genes from the ESX-1 locus[6].

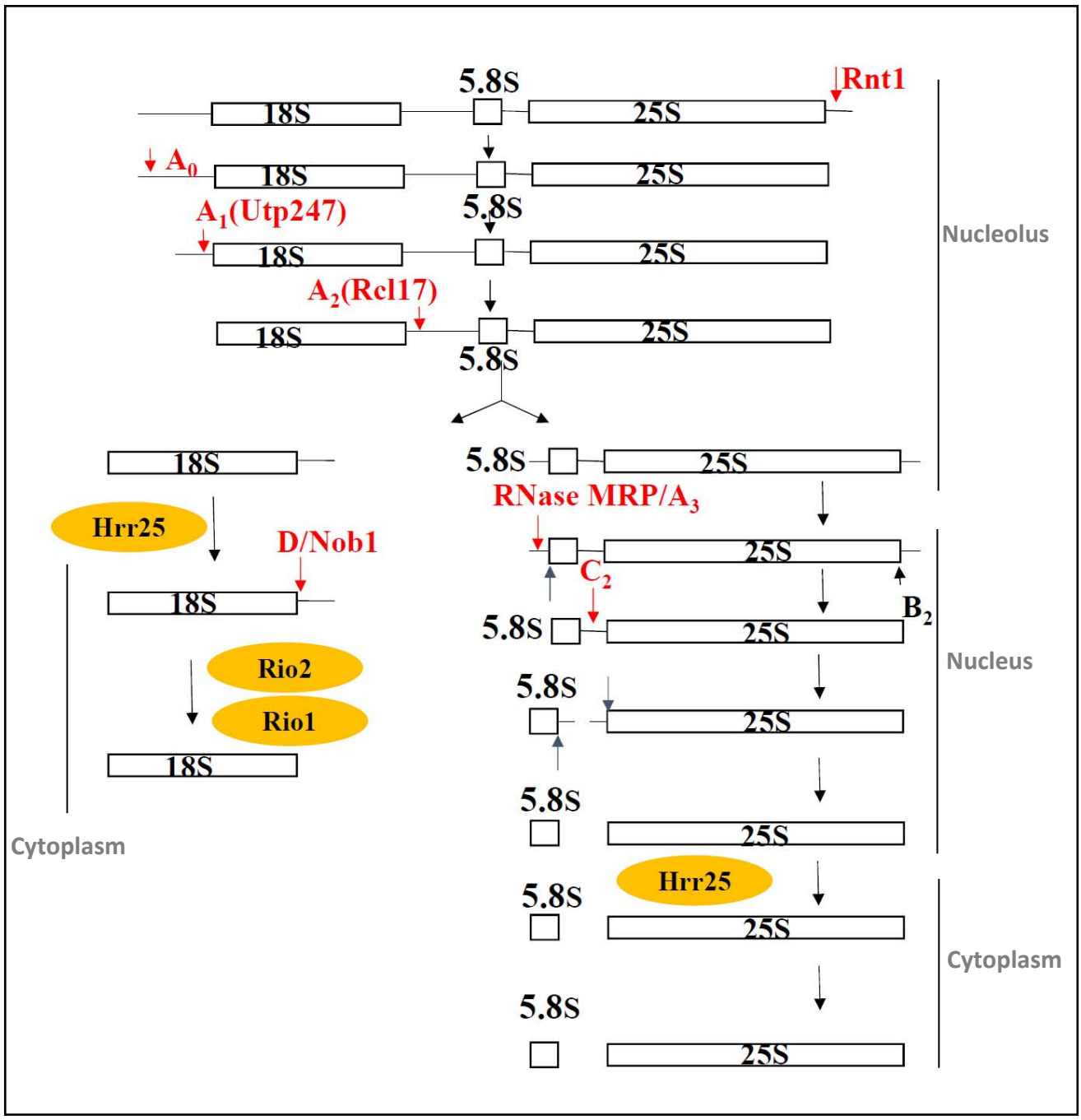

Figure 1.6 The representation of rRNA maturation: The 90S complex is cleaved on the 35S pre-rRNA at processing sites A_0, A_1 and A_2, within the large 90S complex, to produce the pre-40S and the pre-60S particles. The pre-40s particle is rapidly exported to the cytoplasm but the pre-60S particle remains in the nucleolus and undergoes a complex maturation pathway in the nucleoplasm and is subsequently, delivered to the cytoplasm.

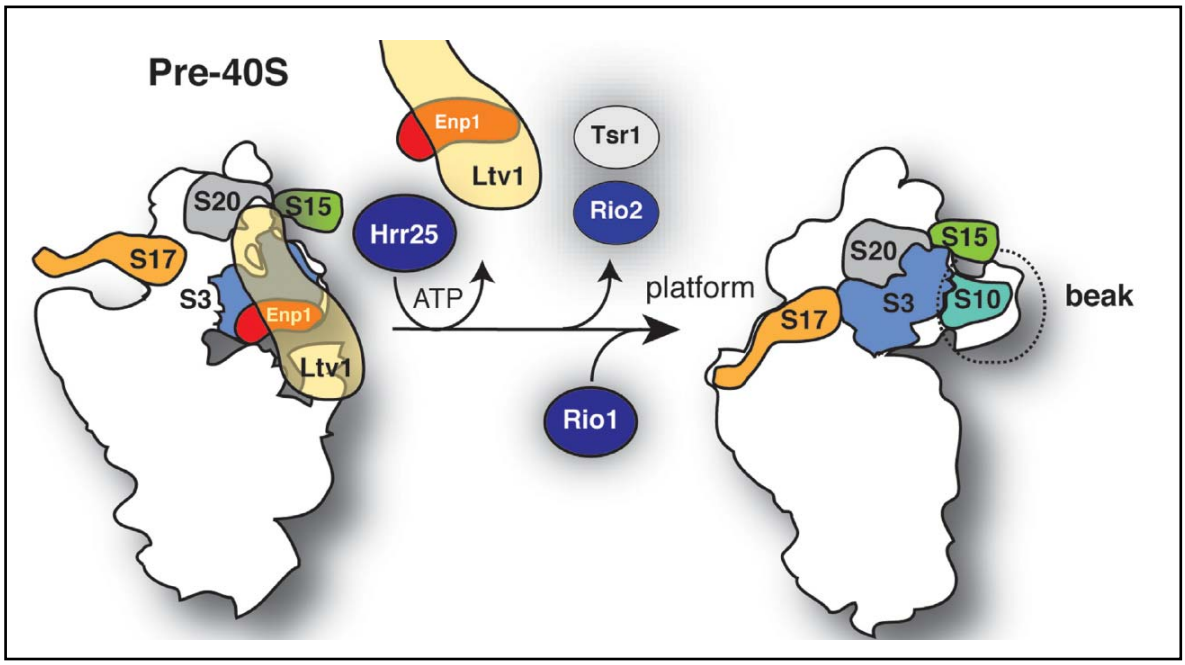

Figure 1.7 Formation of ribosomal beak: In the cytoplasm, the major structural rearrangement that occurs on the pre-40S is the formation of the ribosomal "beak" structure that requires a cascade of phosphorylation and dephosphorylation events. Cryo-EM and biochemical studies have revealed that the formation of the beak is concomitant with the release of assembly factors such as Ltv1 and Enp1[31].

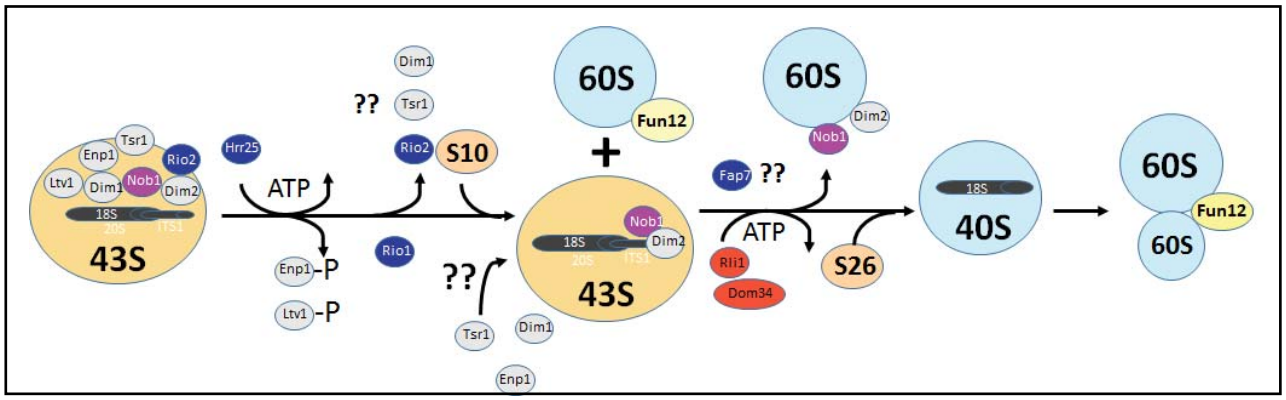

Figure 1.8 Schematic representation of 40S subunit maturation in *Saccharomyces cerevisiae*: Pre-40S particles at different stages of assembly are purified using TAP to obtain the systematic information about cytoplasmic synthetic pathway of maturation. In the cytoplasm, the 20S pre-rRNA is first, methylated at the 3′ end of the 18S rRNA sequence by the function of Dim1.

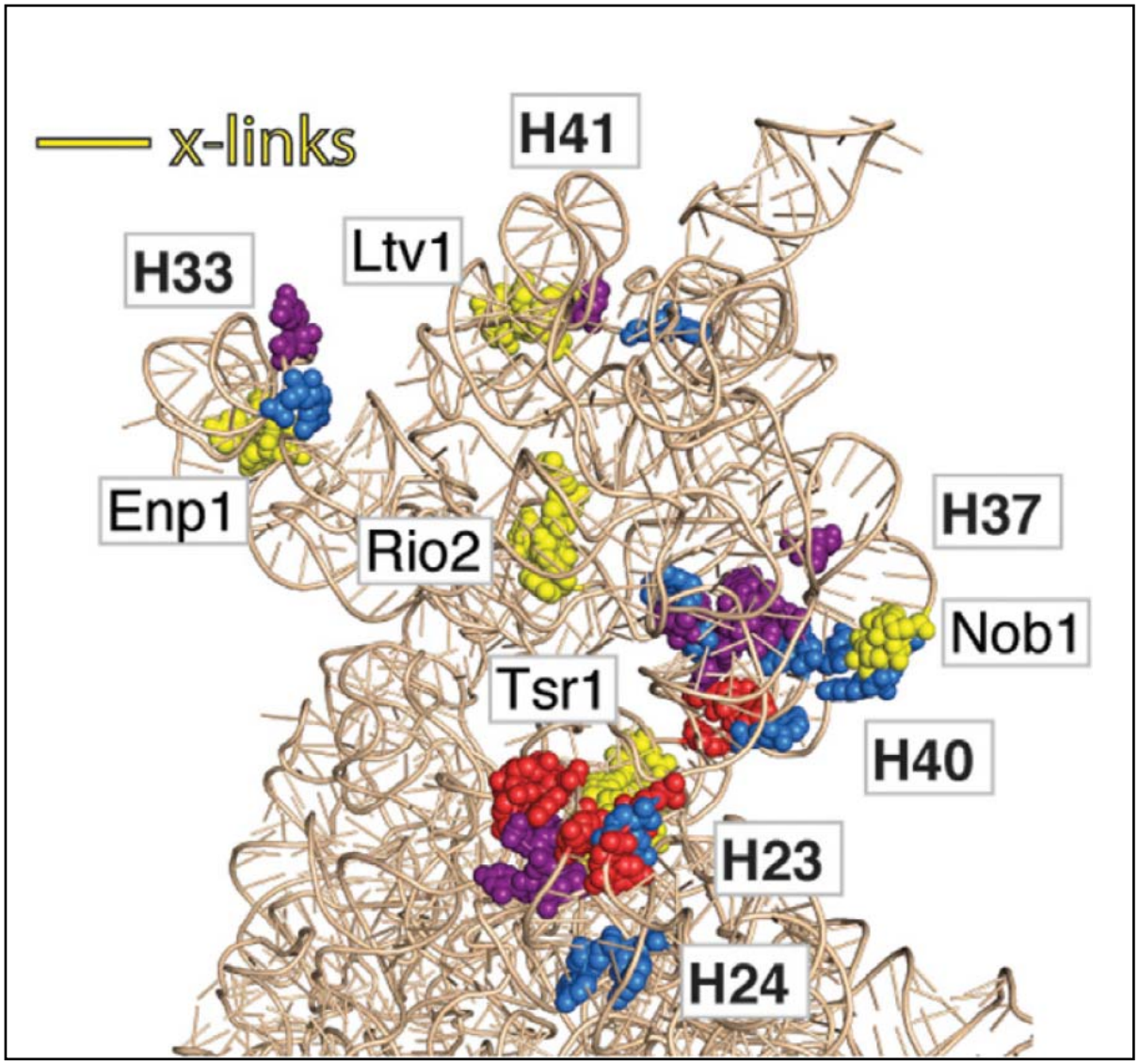

Figure 1.9 Schematic representations of ribosomal protein and rRNA binding sites: In the 40S structure model, the Rio2-binding site is located near the Rpss15. Leger-Silvestre et al have reported that the pre-40S ribosome that lacks Rps15 fail to bind Rio2 and be exported to cytoplasm. In bacteria, the Rps18 and Rps16 homologues (S13 and S9) are phosphorylated at serine and threonine residue[38].

Chapter 2: Purification and Crystallization of EspB

Protein crystallography dates back to 1934 when J.D. Benal and Dorothy Crowfoot obtained crystals and diffraction of hydrated pepsin crystals[52]. The first protein crystal structure was later solved for the sperm whale myoglobin in the late 1950s by Max Perutz and Sir John Cowdey Kendrew for which they both received the Nobel Prize in chemistry in 1962[53]. Since then over 108,395 biological macromolecular structures have been solved and deposited to the protein data bank (PDB). Along with the increased number of solved protein structures, x-ray crystallography methods have also advanced with better detectors and improved software for structure determination.

However, one obstacle that crystallographers face is the phase problem that arises during data collection when the crystal diffracts x-ray beams that are recorded as reflections. These reflections are treated as complicated waves and the ultimate goal of a crystallographer is to obtain an electron density map that reveals the molecular shape of the protein. In order to obtain the electron density map the amplitudes, frequencies and phases have to be determined. Only two of these parameters are directly acquired during data collection, the amplitude and the frequencies. Unfortunately, the phases cannot be determined directly and methods for acquiring phases have been developed that include molecular replacement and single wavelength anomalous dispersion (SAD).

Molecular replacement uses phases from a previously solved crystal structure as initial phases for the new protein. For molecular replacement to be successful, the sequence identity between the new protein and the known protein, also known as the search model, has to be at least 30%. Unlike molecular replacement, SAD phasing relies on heavy atoms to obtain phases for the new protein. The heavy atoms can be introduced by either soaking in ionic solutions or as

modified amino acids during expression. The most common heavy atom is selenium and it is incorporated into the protein during expression in media that has methionine replaced with selenomethionine (Se-Met). Once data is collected, the heavy atoms are located by Patterson maps and the initial phases are calculated and then improved by density modification.

The T7SS and the ESX-1 secretion pathway are not fully understood and obtaining the crystal structure of EspB could provide information regarding the T7SS and the transport mechanism of this protein[16]. We could not use molecular replacement for this protein because there was no known search model for EspB. Therefore, the common heavy atom, selenium, is incorporated into the protein during expression in media that has methionine replaced with selenomethionine (Se-Met). The crystal structure of EspB containing the selenium in the structure was obtained in our lab, previously. In this chapter, I will explain how both of the Full length EspB and the EspB292 were purified in the presence and absence of selenium for the crystallization and characterization studies that are described in chapters four.

2.1 Expression of EspB292

2.1.1 Expression of EspB in the Presence of Selenium

Two constructs of EspB from *M. Marinum*[10], EspB residues 1-292 and EspB residues 1-346 has been expressed, purified and crystallized by my colleague Irene Njeri. For information on this project the reader is referred Irene's dissertation. These constructs were designed based on a limited proteolysis results that identified EspB292 as a stable fragment. Both of these constructs have been prepared and expressed in 1L of seleno-methionine minimal media (Shanghai, Medilon Inc) with 34 μg/mL chloramphenicol and 100 μg/mL ampicillin, to generate Se-Met protein. The cells are then grown at 37°C while shaking at 250 rpm and as soon as the

OD_{600} of the culture has reached 1.2, 1 mM of Isopropyl β-D-1-thiogalactopyranoside (IPTG) is added to the media for induction and expression is preceded at 18°C for 24 hours. The

2.1.2 Expression of EspB in the Absence of Selenium

In a similar procedure, I purified both EspB292 and Full-length EspB (FL-EspB) proteins in the absence of Se-Met. The constructs from M. Marinum were designed based on the limited proteolysis results and was further expressed, purified and crystallized. EspB plasmids with N-terminal histidine tags were transformed to Rosetta^{TM}(DE3) pLysS E.coli cells (Novagen) and were expressed with 34 μg/mL chloramphenicol and 100 μg/mL ampicillin (**Figure 2.1**).

2.2 Purification of EspB292 and FL-EspB

The media was prepared using 4L flasks that were previously washed with soap and distilled water. Each flask was then filled with 1 L distilled water and 20 g LB Broth media without selenium. These solutions were autoclaved for 1 hour and were used to grow bacterial cells in large scale. The overnight culture in the same way was prepared using 500 ml flask that was filled with 100 ml water and 2 g of LB Broth media. To the autoclaved media in the small flask, 100 μg/mL carbenicillin and 34 μg/mL chloramphenicol were added. The frozen bacterial cells were kept at -80ºC as the glycerol stocks. The cells from glycerol stock (in -80ºC freezer) were added to the small flask. This step was done very quickly so that the glycerol stock wouldn't thaw. After adding the cells, the glycerol stock was stored back in -80ºC freezer and the flask was covered with aluminum foil and was grown in 37ºC shaker overnight.

The next day 100 μg/mL carbenicillin and 34 μg/mL chloramphenicol, that were freshly prepared, were added to the large 4 L flasks. The flasks were kept in 37ºC for 30 minutes. Then, 10ml of overnight culture was added to each 4L flask using a clean disposable pipette and the cells were grown in the new media in 4L flasks at 37ºC until the OD_{600} reached 0.6. After almost

three hours the flasks were taken to the 18ºC shaker to cool down and after 10 minutes 1 ml IPTG was added to each flask and the flasks were left at 18ºC to shake overnight. The IPTG was added to the media for induction and expression that proceeded at 18°C for 24 hours.

The cells were centrifuged at 5,000 rpm and 4°C for 20 minutes using a Sorvall centrifuge (Dupont) with a SLC 4000 rotor. The bacterial pellets could be stored at -80ºC to be resuspended later. In my experiment, right after the centrifugation the bacterial cells were resuspended in lysis buffer containing 50 mM Tris pH 8.0, 200 mM NaCl, 0.05% β-mercaptoethanol, 10% glycerol and one tablet of protease inhibitor. The resuspended bacterial cell pellet was subsequently, lyzed using 3 ml bugbuster followed by French press and the lysate was centrifuged at 4°C and 15,000 rpm for 50 minutes using a Beckman Coulter ultracentrifuge with a 45 Ti rotor.

The supernatant was next filtered and loaded onto a pre-equilibrated 5 mL HisTrap HP column (GE healthcare) that was attached to a Fast Protein Liquid Chromatography (FPLC) system at the flowrate of 2.5 ml/min. The Histrap column was previously, washed and equilibrated with several rounds of buffer A and buffer B.

Buffer A was prepared using 25mM Tris pH 8, 200mM NaCl and 0.05% BME and 10% glycerol. Buffer B was prepared using 250mM Imidazole, 25mM Tris pH 8 and 200mM NaCl, 10% glycerol and 0.05% BME (**Figure 2.2**).

After the lysate was loaded on the column, it was washed with 10 column volumes of lysis buffer (Buffer A) to remove any unbound protein. Subsequently, the bound protein was subjected to an imidazole gradient of 20 mM to 250 mM and EspB eluted between 100 mM and 150 mM of imidazole. Partially purified fractions were pooled together and transferred to a dialysis membrane to which 1 mg/mL of *Tobacco Etch virus* (TEV) protease was added to

cleave the histidine tag. EspB292 was dialyzed overnight at 4°C against 20 mM Tris pH 8.0, 50 mM NaCl, 0.05% β-mercaptoethanol and 10% glycerol. After cleavage the protein was loaded onto a pre-equilibrated 5 mL HisTrap HP column (GE healthcare) and was washed with the same buffer that was used for dialysis. The cleaved protein was supposed to be collected in the flow through. The cleaved protein was then concentrated to 5ml and was further purified by size exclusion chromatography using a Superdex 200 column (GE healthcare) equilibrated in 10 mM HEPES, 50 mM NaCl and 0.05% β-mercaptoethanol at pH 7.5. All purification steps were monitored by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS PAGE) for purity (**Figure 2.3**).

2.3 Crystallization of EspB292 and FL-EspB

Purified EspB292 was concentrated to 25mg/mL and was subjected to robotic sparse matrix screening using Phoenix Liquid Handing System (Art Robbins). The procedure is again followed from Irene's dissertation and commercially available matrix screens, Index (Hampton), Cyros (Qiagen), PEGs suite (Qiagen), Wizard I, II and III (Emerald Biosystems) were used for initial crystal screening in 96 well plates with three different protein to well solution ratios (1:2, 1:1, 2:1). Crystals that were observed during the initial screening were small and didn't generate a good diffraction quality crystal. For Se-Met292, three different crystals were produced in our lab by Irene. The crystal with diamond-shaped in Wizard I and III suite grew in 1.2 M NaH_2PO_4, 0.8 M K_2HPO_4, 0.1 M CAPS pH 10.5, 0.2 M Li_2SO_4 (**Figure 2.4**). Diffraction quality crystals were flash frozen in their respective well conditions with 20 % (v/v) ethylene glycol as the cryo-protectant.

Following the similar procedure, I tried the crystallization of EspB292 and FL-EspB in presence and absence of copper ions. Although several crystals were observed, none of them diffracted to a high quality and resolution.

Many factors influence the likelihood of crystallization of a protein sample. Some of these factors include protein purity, pH, concentration, temperature, precipitants and additives. The more homogeneous the protein solution is, the more likely that it will crystallize. Typically, protein samples above 97% purity are considered suitable for crystallization, although high purity is neither necessary nor sufficient for crystallization. Solution pH can be very important and in extreme cases can result in different packing orientations. Buffers, such as Tris-HCl, are often necessary for the maintenance of a particular pH, Precipitants, such as ammonium sulfate or polyethylene glycol, are usually used to promote the formation of protein crystals. In the case of EspB protein one reason for the low quality of crystals is the degradation of this protein over time. During the incubation of the crystals at 20ºC the solution of EspB could get less homogeneous because of the probable cleavages at different sites. This was more observed for the full length EspB. Therefore, only selenium methionine crystals were good enough for diffraction.

2.4 Data Collection and Structure Determination

Data that has been collected at 100K at the NE-CAT beamline at the Advanced Photon Source (APS), Argonne, IL, USA were indexed, integrated and scaled. The diamond-shaped crystals of EspB292 grown with the selenium diffracted to 2.5 Å. The crystals belong to P21 space group and the unit cell dimensions are a= 173.448 Å, b= 81.703 Å, c= 267.467 Å and α=90.000, β=108.770, γ=90.000 (**Table 2.1-2.2**).

2.5 Overview of the Structure of EspB

Based on the crystal structure the cleaved EspB292 has 16 monomers in the asymmetric unit (molecules A-P) that are arranged into a set of four homotetramers (**Figure 2.5**). Two sets of these tetramers (tetramers 1 and 3; tetramers 2 and 4) are arranged parallel to each other. Contact is observed between the C-terminal loop of one tetramer and the C-lobe helices of the adjacent tetramer. Every homo-tetramer has two monomers (monomers A and B and monomers C and D) that are arranged anti-parallel to each other creating an interacting interface and a central helix bundle (**Figure 2.6-2.7**). Contact is also observed between the C-lobe helices of two monomers within the tetramer. For instance monomer A interacts with monomer D and monomer B interacts with C via the C-lobe helices (**Figure 2.8**). The monomers have regions on the N- and C-terminus that lack electron density due to disorder and flexibility on these ends.

EspB292 is composed of long, antiparallel coiled-coil helices. EspB292 can form hetero dimer with Rv3879c and be translocated through the inner membrane. Based on the crystal structure we can also observe that EspB can adopt a homodimer fold.

2.6 EspB can Adopt a Homodimer Fold

The helix-turn-helix domains of CFP10 and the N-terminal domain of EspB292 residues 1–92 are roughly superimposable, each with a characteristic YxxxD motif at the turn. EspB292, residues 135–226, and ESAT-6, residues 2–104, are also roughly superimposable, each with a WxG motif at the turn of the second helix-turn-helix subdomain (**Figure 2.9**). In EspB292, these two subdomains form an extensive interface of non-covalent interactions, and the linker connecting them adopts an extended conformation that anchors loosely to the helical bundle through van der Waals interactions.

EspB292 possess features that are not found in ESAT-6 and CFP10. Notably, helix 3, helix 6, and helix 7 of EspB292 create a ‘‘helical tip’’ that is rich in solvent-exposed hydrophobic/aromatic/acidic residues (**Figure 2.10**). The tip ends in an extended polyproline stretch (residues 270–288) that sandwiches between the subdomains through hydrophobic interactions provided by prolines and aliphatic side chains.

2.7 EspB can Target the EccCb1 ATPase

EspB has a bipartite signal sequence that targets the EccCb1 ATPase. The EspB292 crystal structure reveals the unique relative disposition of multiple characteristic ESX sequence motifs, including a structured view of the helical export arm. In this work, we define the export arm as EspB292 residues 78–93, corresponding to the appendage that is dynamic in the structure and lacking electron density. The prevalent view is that residues from the export arm align on the same helical face upon binding the EccCb1 ATPase, targeting ESX secreted substrates for secretion

2.8 Secretion of EspB/Rv3879c Follows Pairwise Dependency

The three dimensional structure of a protein provides insights into its probable function. Therefore, proteins with similar functions often tend to adopt similar folds even when they have very low sequence similarity. The results obtained from the x-ray crystallography indicate novel functional roles for the secreted proteins Rv3881c (EspB). The 3D fold structure of EspB292 shows a helix-turn-helix fold at the N-terminal of the protein and the C terminal end of EspB292 has a WEG tripeptide motif at the position 176–178 that is similar to the WXG motif observed for the ESAT-6 like family of proteins. Based on previous results, the N-terminal region of Rv3879c displays helical regions followed by coiled-coil regions, similar to fibrinogen fold. These observations suggest that the N-terminal regions of EspB and Rv3879c can form a four-

helix-bundle similar to CFP-10/ESAT-6 pair. Interestingly, the 3D fold analysis of EspC protein also adopts a similar fold to that of ESAT-6 and EspB and the N-terminal region of EspA exhibited WXG motif. With regards to the secretory proteins of the ESX-1 secretion pathway, the current analysis indicates that the secretory protein pairs have a preference to adopt a four-helix-bundle structure. While the CFP-10/ESAT-6 and EspA/EspC pairs were previously reported to have such a structure, the present analysis suggests that the other secretory protein pairs EspB/Rv3879c may also adopt similar structures.

2.9 Inner Membrane Translocation of EspB/Rv3879c Pair

Previous studies have provided experimental evidences that Rv3868 interacts with CFP-10/ESAT-6 and EspA/EspC pairs and helps in their multimerization. This multimeric protein assembly as well as EspB/Rv3879c is recognized by Rv3870 that is a member of Ftsk/SpoIIIE family of DNA translocase proteins. Rv3870 contains two transmembrane helices and a single FtsK ATPase domain and is known to interact with another cytoplasmic FtsK/SpoIIIE family protein Rv3871 that contains two FtsK ATPase domains. The α and β units of the FtsK domain containing proteins are known to multimerize to produce a hexameric ring with an inner diameter of 30Å. During cell division, this hexameric multimer acts as a DNA pump which translocates double stranded DNA with diameter of 20 Å. The α and β units domains of Rv3870/Rv3871 pairs also forms a hexameric ring, similar to that formed by a hexamer of FtsK proteins. Structural analysis of CFP-10/ESAT-6 four-helix-bundle diameter has previously shown to be approximately 20–23 Å. Our findings suggest that each EspB292 monomer is similarly alpha helical and is roughly shaped like a cylinder with 103.6 Å length and a diameter of 16.1 Å that would perfectly fit and translocate through the hexameric ring of Rv3870/Rv3871 pairs. This study suggest that, the FtsK family proteins (Rv3870 and Rv3871) recognize and translocate the

four-helix substrates similar to that used by FtsK family proteins for the translocation of double stranded DNA. Therefore, the Rv3870/Rv3871 hexameric ring of T7SS may be a good example of machinery that has evolved from a specialized double stranded DNA translocator to a protein transporter.

The consecutive step after recognition of the secretory proteins by Rv3870/Rv3871 ring is translocation of these proteins through the inner membrane pore of Rv3877. The diameter of the inner pore considering a compact arrangement of the helices is determined to be approximately 26 Å or larger upon incorporating the side-chain lengths and interactions. Thus, this data also suggests that the substrate pairs (four-helix-bundles) of the ESX-1 secretion pathway would be able to fit into and be translocated along the inner pore of Rv3877.

2.10 Secretion Regulation of CFP-10/ESAT-6

Once EspB pairs up with Rv3879c, it co-secretes with CFP-10/ESAT-6 and EspA/EspC through Rv3877 inner pore to the periplasm-like space. These multimers thereby, reach MycP1 protease which is anchored to the Rv3877 by a C-terminal transmembrane helix, and is located in the periplasm-like space.

Therein, MycP1 plays a dual role in regulating the secretion activity of ESX-1. MycP1 interacts with Rv3877 and opens the pore to enable protein secretion. Thus, the knock-out of MycP1 has shown to completely abolish secretion. Nonetheless, its protease activity negatively regulates secretion via EspB. The presence of a mutant MycP1 without any protease activity has shown to increase the secretion of CFP-10/ESAT-6 and accumulation of EspB in the periplasm-like space. These findings suggest a probable regulatory role of EspB cleavage by this serine protease (MycP1) in the periplasm-like space since the cleavage of EspB by MycP1 leads to controlled secretion of both the CFP-10/ESAT-6 and EspA/EspC multimers.

2.11 Summary

The crystal structure of EspB292 has 16 monomers in the asymmetric unit which is composed of long, antiparallel coiled-coil helices. Based on the structure EspB292 can interact with other proteins and form heterodimers and interact with itself and form homodimer. To form heterodimers the helix-turn-helix domains of the N-terminal domain of EspB292 residues 1–92 are used. EspB292, residues 135–226, and ESAT-6, residues 2–104, are also roughly superimposable, each with a WxG motif at the turn of the second helix-turn-helix subdomain. The WxG motif is required for the formation of the dimers. It's been shown in TB that at least one of the monomers in a dimer should have the WxG sequence.

EspB292 possess features that are not found in ESAT-6 and CFP10. Notably, helix 3, helix 6, and helix 7 of EspB292 create a ''helical tip'' that is rich in solvent-exposed hydrophobic/aromatic/acidic residues. The tip ends in an extended polyproline stretch (residues 270–288) that sandwiches between the subdomains through hydrophobic interactions provided by prolines and aliphatic side chains.

The crystal structural information also reveals that EspB has a bipartite signal sequence and structure that targets the EccCb1 ATPase. The EspB292 crystal structure reveals the unique relative disposition of multiple characteristic ESX sequence motifs, including a structured view of the helical export arm. In this work, we define the export arm as EspB292 residues 78–93, corresponding to the appendage that is dynamic in the structure and lacking electron density.

Previous studies have provided experimental evidences that Rv3868 interacts with CFP-10/ESAT-6 and EspA/EspC pairs and helps in their multimerization[54]. This multimeric protein assembly as well as EspB/Rv3879c is recognized by Rv3870 that is a member of Ftsk/SpoIIIE family of DNA translocase proteins. These FtsK proteins produce a hexameric ring with an inner

diameter of 30Å. Structural analysis of CFP-10/ESAT-6 four-helix-bundle diameter has previously shown to be approximately 20–23 Å. Our findings suggest that each EspB292 monomer is similarly alpha helical and is roughly shaped like a cylinder with 103.6 Å length and a diameter of 16.1 Å that would perfectly fit and translocate through the hexameric ring of Rv3870/Rv3871 pairs.

Once EspB pairs up with Rv3879c, it co-secretes with CFP-10/ESAT-6 and EspA/EspC through Rv3877 inner pore to the periplasm-like space. These multimers thereby, reach MycP1 protease which is anchored to the Rv3877 by a C-terminal transmembrane helix, and is located in the periplasm-like space.

EspB292 construct:

N´---6x His Tag--- TEV Site--- EspB292---C´

Full Length EspB:

N´---6x His Tag--- TEV Site--- FL-EspB---C´

Figure 2.1 Schematic representation of EspB constructs: EspB constructs are designed with an N-terminal 6x His tag followed by a TEV cleavage site.

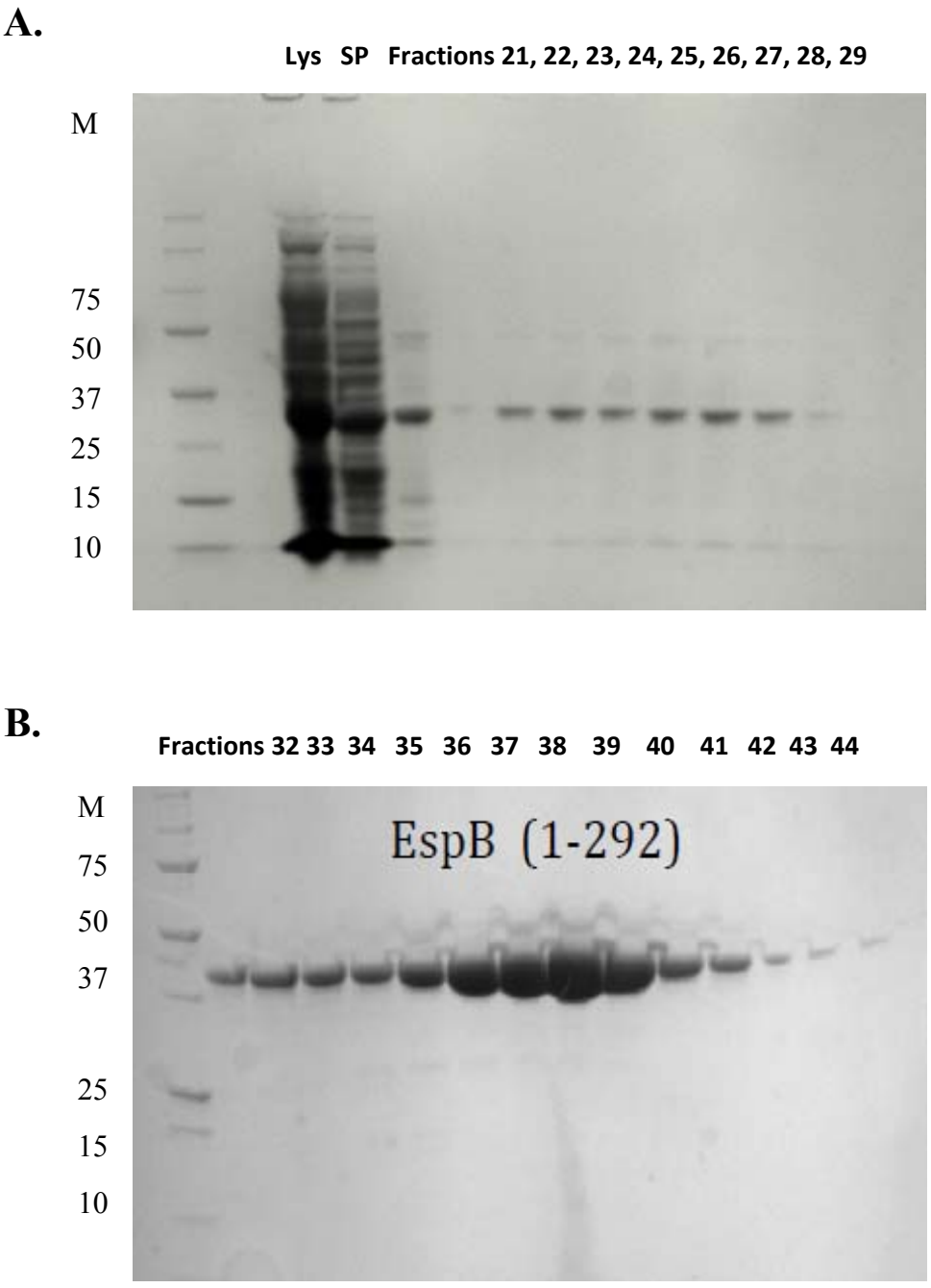

Figure 2.2 Preparative gels of EspB292 after first His-trap column (A) and size exclusion chromatography (B): Lane M is the protein ladder. Gel of EspB292 displays pure protein at the expected molecular weight of 31.8kDa. lys means Lysate, and SP is Supernatant.

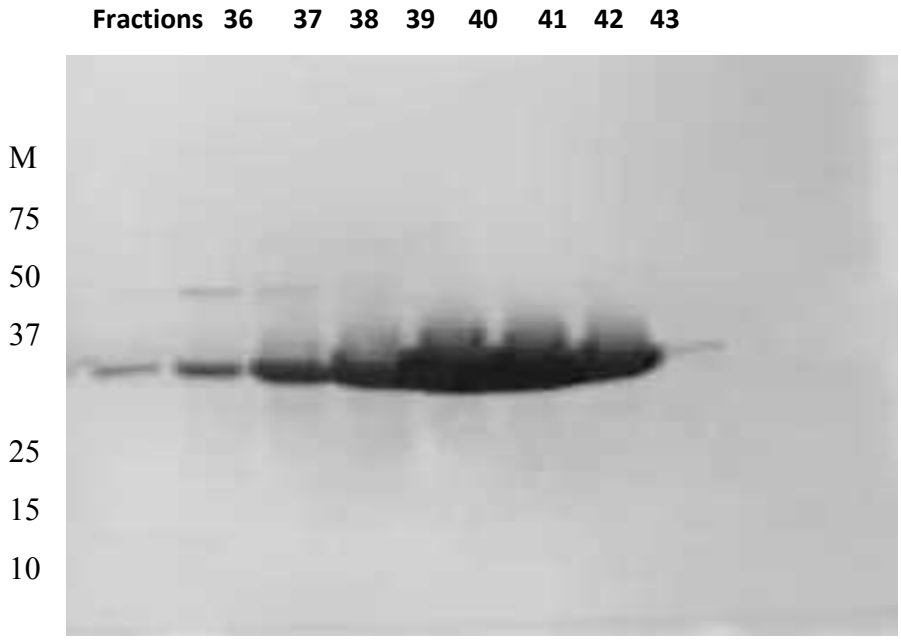

Figure 2.3 Preparative gels of FL-EspB after size exclusion chromatography: Lane M is the protein ladder. Gel of EspB292 displays pure protein at the expected molecular weight of 37.8kDa.

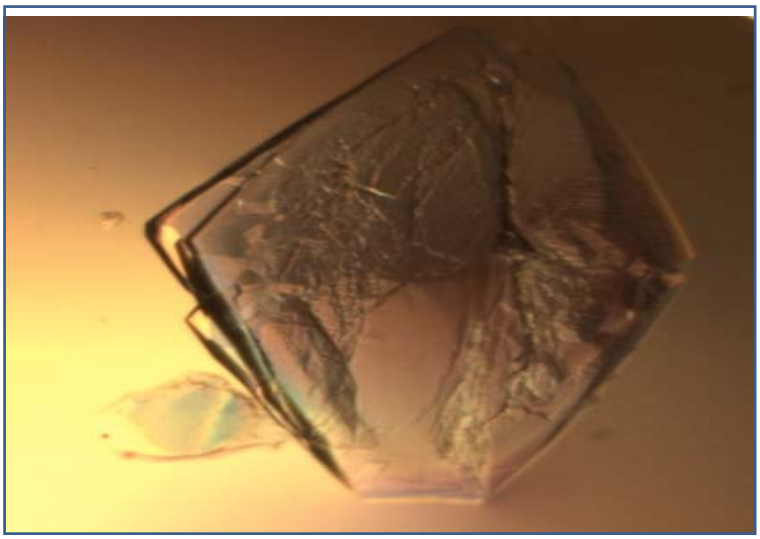

Figure 2.4 EspB292 crystals: Diamond-shaped crystals at pH 8.5 that diffracted to 2.5Å.

Table 2.1 The Se-Met EspB292 Scalelog file. The crystal diffracted to 2.5Å.

Shell Lower limit	Shell Upper limit	Average Redundancy Per Shell
50.00	5.38	3.9
5.38	4.27	3.8
4.27	3.73	3.9
3.73	3.39	3.9
3.39	3.15	3.9
3.15	2.96	3.9
2.96	2.82	3.8
2.82	2.69	3.8
2.69	2.59	3.5
2.59	2.50	3.0
All hkl		3.7

Shell Lower limit	Shell Upper limit	I/Sigma in resolution shells: % of of reflections with I / Sigma less than 0	1	2	3	5	10	20	>20	total
50.00	5.38	0.2	0.9	2.0	3.2	5.9	16.0	99.9	0.0	99.9
5.38	4.27	0.6	1.7	3.4	5.2	9.3	24.2	100.0	0.0	100.0
4.27	3.73	1.3	3.2	5.7	8.4	14.3	33.4	100.0	0.0	100.0
3.73	3.39	2.2	5.6	10.1	14.8	23.8	48.4	96.8	3.2	100.0
3.39	3.15	3.9	10.0	17.9	25.4	39.5	67.9	98.3	1.7	100.0
3.15	2.96	6.6	16.6	28.3	39.0	56.3	82.8	98.8	1.2	100.0
2.96	2.82	9.1	22.8	38.4	51.0	68.7	90.0	99.3	0.6	100.0
2.82	2.69	12.6	30.7	49.6	63.5	80.3	95.4	99.7	0.2	99.9
2.69	2.59	16.0	39.5	61.1	74.5	88.0	97.5	99.2	0.0	99.2
2.59	2.50	19.2	46.4	67.8	79.4	89.3	94.8	95.6	0.0	95.6
All hkl		7.2	17.8	28.5	36.5	47.6	65.0	98.8	0.7	99.5

Shell limit	Lower Angstrom	Upper Angstrom	Average I	Average error	Average stat.	Norm. Chi**2	Linear R-fac	Square R-fac
50.00	5.38		76.9	4.5	2.0	0.985	0.062	0.071
5.38	4.27		59.1	3.8	2.0	1.316	0.092	0.104
4.27	3.73		51.4	3.6	2.2	1.472	0.111	0.118
3.73	3.39		34.9	2.8	2.2	1.405	0.131	0.133
3.39	3.15		21.6	2.4	2.2	1.128	0.172	0.165
3.15	2.96		13.7	2.4	2.3	1.002	0.247	0.237
2.96	2.82		10.4	2.5	2.4	0.947	0.330	0.304
2.82	2.69		7.4	2.6	2.6	0.887	0.452	0.413
2.69	2.59		5.5	2.9	2.9	0.868	0.605	0.559
2.59	2.50		4.4	3.6	3.5	0.850	0.714	0.649
All reflections			28.6	3.1	2.4	1.098	0.142	0.110

Table 2.2 Se-Met EspB292 data collection and refinement statistics.

Data collection	EspB (1-292)
Space group	$P2_1$
Cell dimensions	
a, *b*, *c* (Å)	173.448,81.703,267.467
α, β, γ (°)	90, 108.77, 90
Molecules/Asym. unit	16
Wavelength (Å)	0.97921
Resolution (Å)	50-2.50
R_{sym} (last shell)	0.142 (0.452)
Completeness (%)	99.5 (95.6)
Redundancy	3.7 (3.0)
Refinement:	
Resolution (Å)	43.8-2.50
R_{work}/R_{free} (%)	21.0/24.9
Mean B-factors $(Å)^2$ Residues	49.8
RMSD Bond lengths (Å)	0.009
RMSD Bond angles (°)	1.176
Ramachandran Plot	
Favored	96.3%
Allowed	0.0%

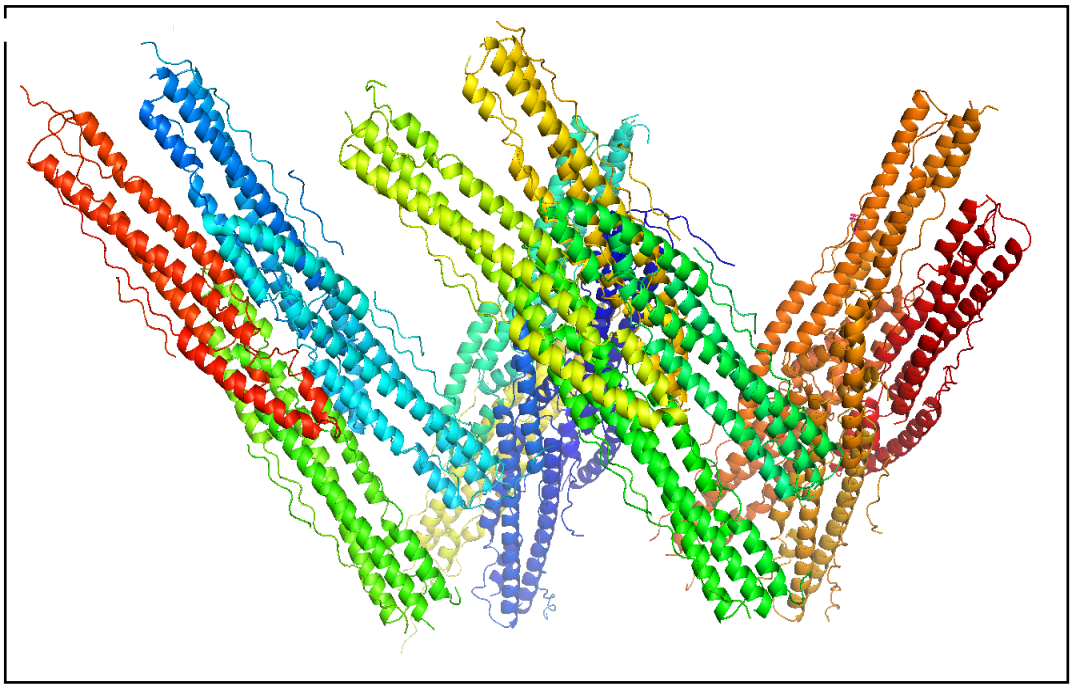

Figure 2.5 Overall structure of EspB292: EspB292 has 16 monomers in the asymmetric unit that are arranged into a set of four homotetramers.

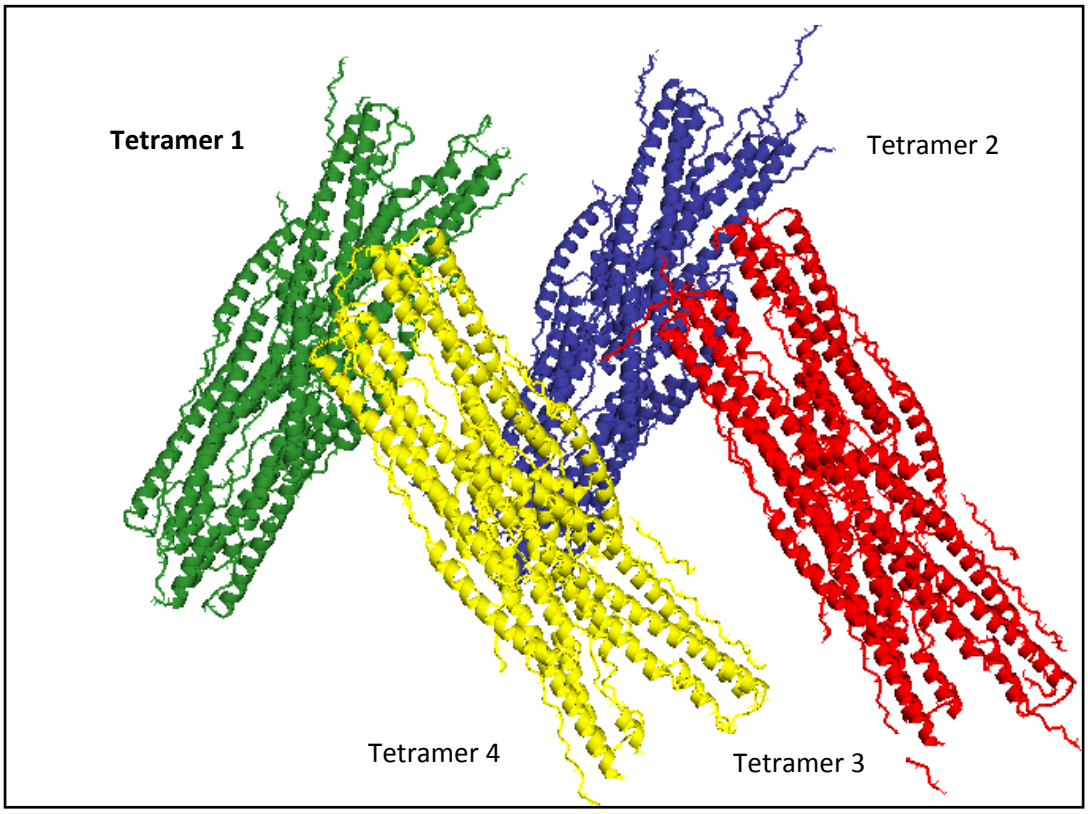

Figure 2.6 Structure of EspB has four homotetramers: Two sets of these tetramers (tetramers 1 and 2; tetramers 3 and 4) are arranged parallel to each other.

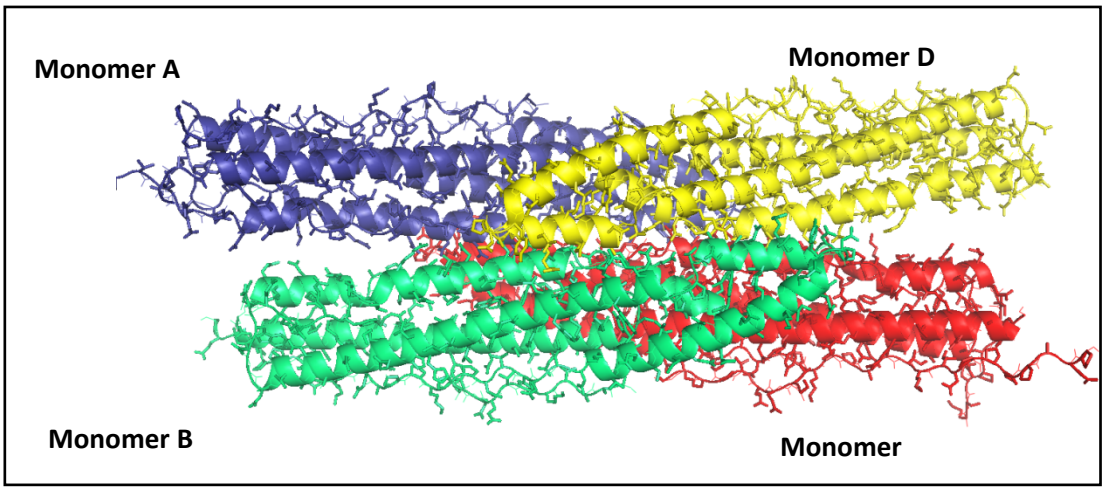

Figure 2.7 Structure of EspB has four homotetramers: Contact is observed between the C-terminal loop of one tetramer and the C-lobe helices of the adjacent tetramer.

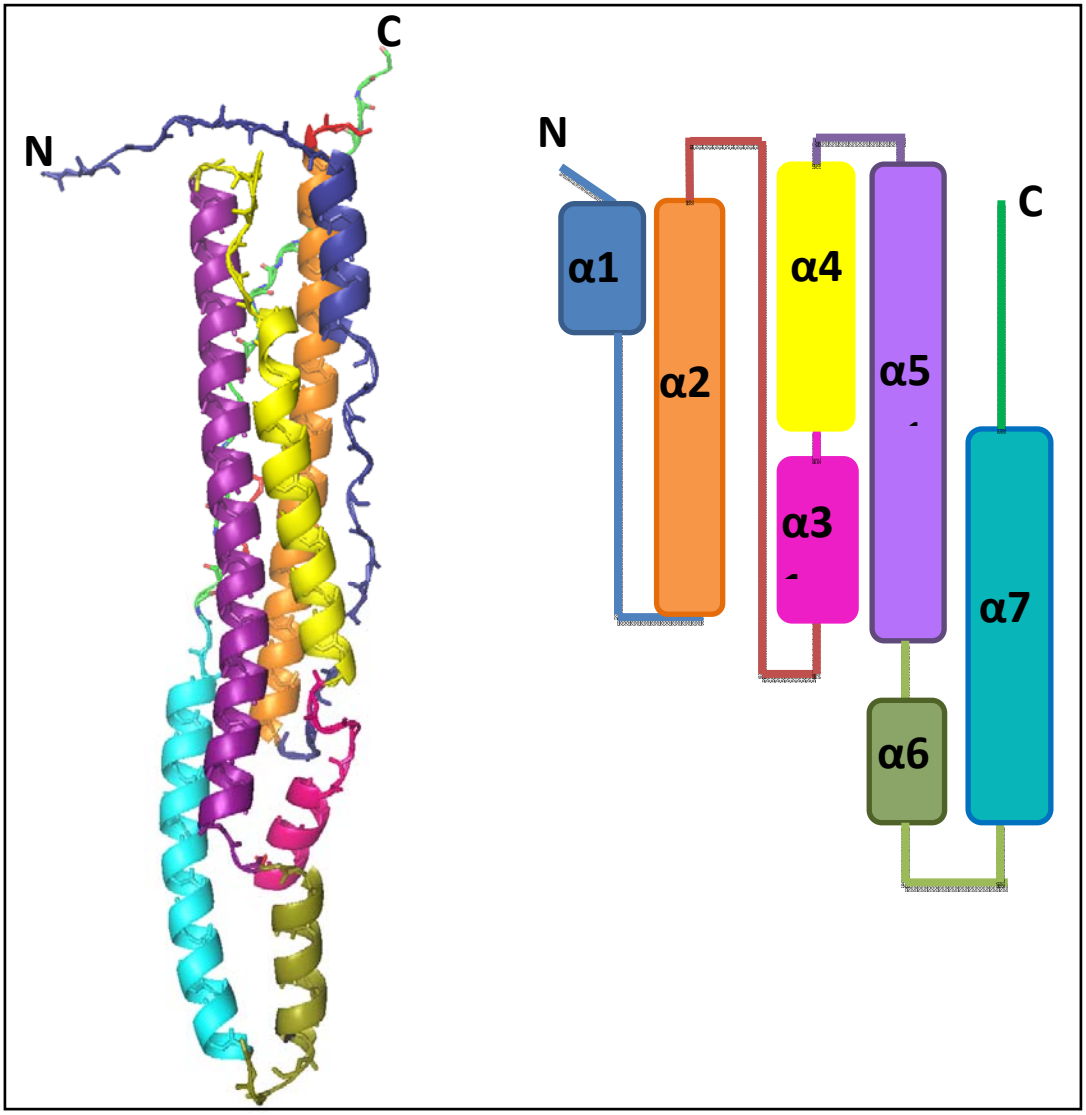

Figure 2.8 Overall monomeric structure of EspB: The monomer is mostly helical and cylindrical in overall shape.

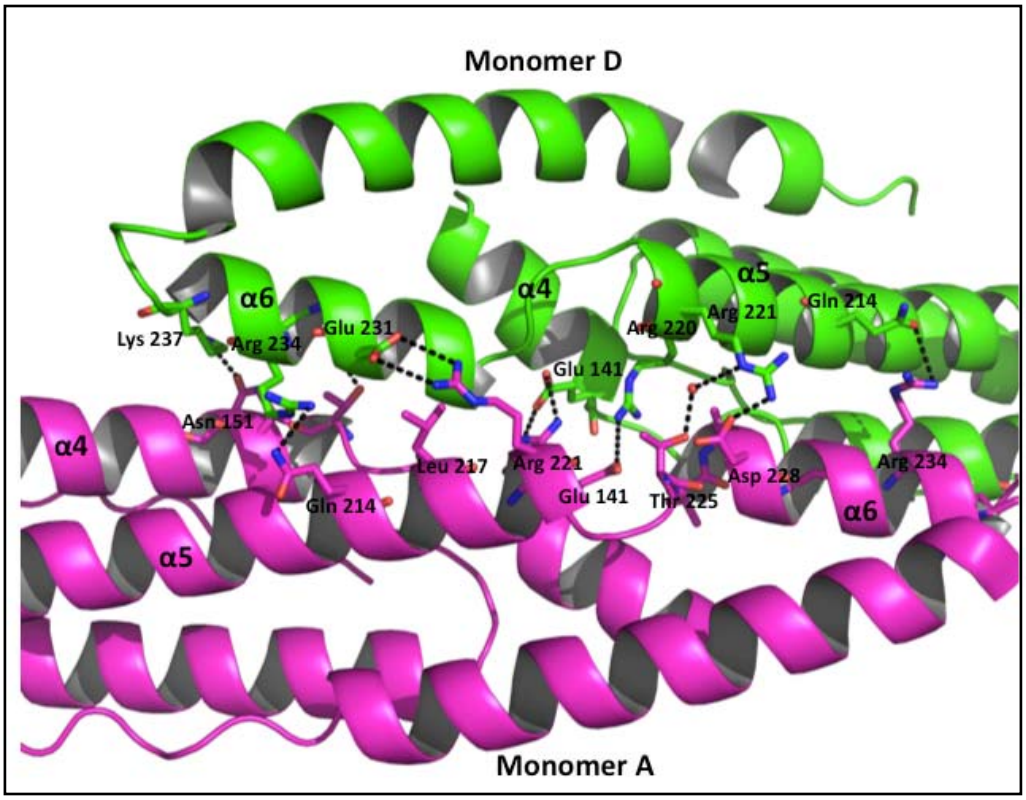

Figure 2.9 Interface between monomers A and D: Black dashed lines are hydrogen bonds and red spheres are waters. Interface between the C-lobe helices (α4, α5 and α6).

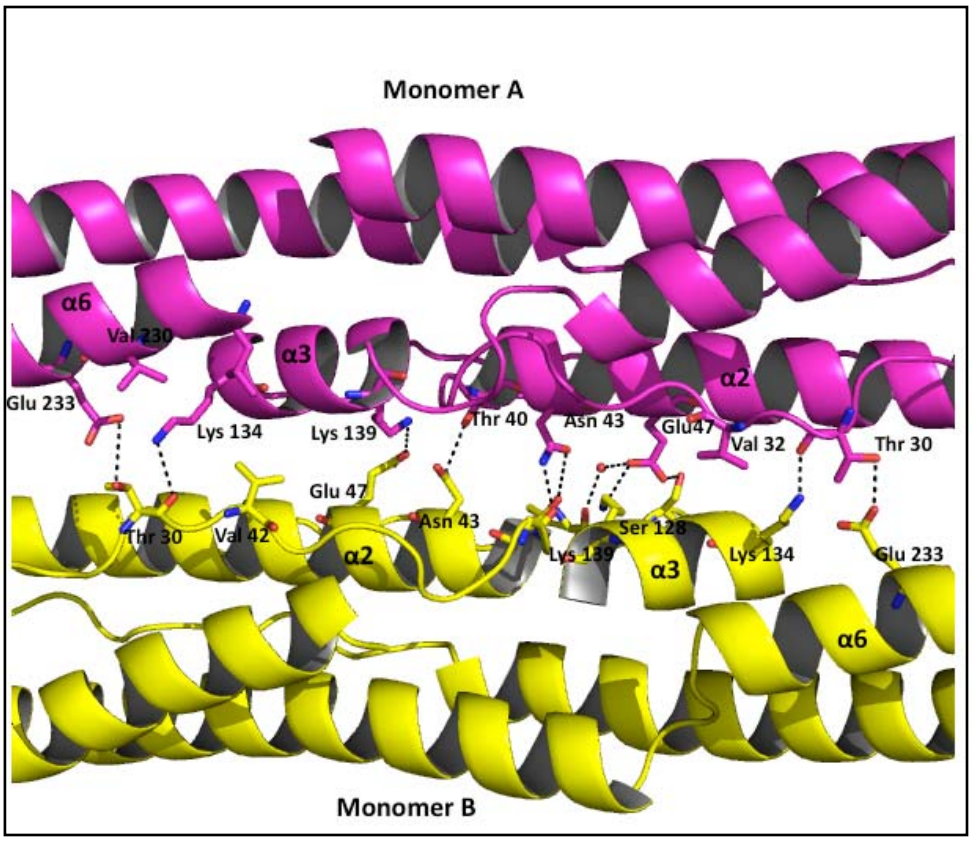

Figure 2.10 Interface between monomers A and B: Black dashed lines are hydrogen bonds and red spheres are waters. Interface between the C-lobe helices (α4, α5 and α6).

Chapter 3: Oligomeric State Characterizations of EspB

3.1 Type VII Secretion System

As discussed earlier in chapter one, Mtb is capable of manipulating its environment by delivering virulence factors during infection[55]. The secretion of variety of virulence factors such as ESAT-6 and CFP10 proteins evade, modulate and control host immune responses and creates a favorable environment for the intracellular survival and replication of Mtb[56]. To penetrate through the highly protective, and hydrophobic waxy coating of mycolic acids these virulence factors are exported by a specialized secretion pathway that is called Type VII secretion system (T7SS).

Based on the T7SS, we know that the virulence of Mtb is initialized by phagosomal signals that trigger the transcription and secretion of two pairs EspA/EspC and CFP-10/ESAT-6. Once expressed, Rv3868 assembles these pairs into multimers that are subsequently recognized by Rv3871 and Rv3870. This translocase pore facilitates the ATP-dependent translocation of the substrates through the inner membrane pore protein (Rv3877)[57]. Thereafter, the interaction of the N-terminal gate-like domain of Rv3877 with MycP1, a serine protease, opens the pore enabling translocation of the substrate assembly from cytoplasm to the periplasm-like space[17, 19, 20, 22, 58, 59]. Ultimately, the substrate assembly crosses the mycomembrane through a channel formed by Rv3869, Rv3882c and the four proteins of MCE cluster 1 (Mce1B, Mce1C, Mce1F and Rv0177) (**Figure 3.1**).

3.2 Role of EspB in Type VII Secretion System

In this proposed model we don't know the much about EspB protein and its transport mechanism. EspB, as well as ESAT-6 and CFP-10, is a very important virulence factor. EspB in particular is directly involved in ESX-1 membranolytic function, and strains lacking this secreted

protein are as attenuated for virulence as those lacking the entire secretion apparatus[17, 20, 21, 55, 57, 59, 60]. Knocking out EspB in the related pathogen Mtb abrogates red blood cell hemolysis, eliminates macrophage cytotoxicity, inhibits intracellular growth and phagosome maturation and completely attenuates virulence.

3.3 EspB292 Oligomization in Literature

Based on previous findings, EspB cleavage regulates the controlled secretion of CFP-10/ESAT-6 and EspA/EspC multimers. Our crystallographic studies, furthermore, suggested that upon cleavage EspB aggregates to a set of four homotetramers. Therefore, the controlled secretion of four-helix-bundles might be as a result of the EspB aggregation. It has been shown by Strynadka et al that in light scattering and negative stain electron microscopy (EM), EspB from Mtb has a propensity to oligomerize, revealing the potential for EspB to serve as structural subunits in the construction of cell wall-associated architectures of mycobacteria as its widely hypothesized. To clarify the components that affect oligomerization of the EspB292, we set up experiments that would mimic the phagosomal environment to find out how the protein self–assembly occurs in biologically relevant conditions.

While our structural studies of the N-terminal domain of EspB292 provide insight into the molecular basis of ATPase-mediated export, subsequent analysis of our full length Mtb EspB revealed that EspB has the capability of forming higher order oligomeric states. Immediate clues of EspB oligomerization arose during purification, with protein eluting on a gel filtration column at multiple retention times, suggesting that this protein can self-associate in a concentration-dependent manner.

To more accurately determine the subunit stoichiometry of the higher molecular species, Strynadka et al carried out size-exclusion chromatography coupled to multi-angle light scattering

(SEC-MALS) analysis. The SEC-MALS experiment detected a monodisperse peak of molecular weight 52 kDa (corresponding to mass of a single full length EspB as well. They also confirmed the oligomerization of EspB by analytical ultracentrifugation sedimentation velocity experiments that identified a low molecular weight 2.1S species (44 ±5 kDa) and a high molecular weight species with a sedimentation coefficient of 10.5S (317 ± 41 kDa). To characterize the structural properties of the full length EspB oligomer, they subjected the peak fractions to gel filtration run to negative stain EM analysis. The data shows ring-shaped structures that were evenly distributed across the grid.

3.4 Results and Discussion

3.4.1 EspB292 Oligomization

Many proteins function through protein complex assemblies therefore, studying protein-protein interactions in biologically relevant conditions, can often be a crucial step in understanding the fundamentals of many biological processes. Based on the crystal structure of EspB292 protein, in the non-covalent protein complex assemblies, the individual protein members reside in close proximity of each other. The close proximity of the lysine residues was utilized to identify a proper cross-linker (BS3) with respective spacer arm length (of 11.4Å) that could covalently link two interacting proteins. Hence, the chemical cross-linking experiment was performed to first link the proteins covalently, followed by well-established protocols such as SDS-PAGE for identification of the protein assemblies.

3.4.2 BS3 Cross-Linking Protocol

Immediately before use, a 50mM solution of BS3 (Bis-Sulfosuccinimidyl-suberate [61]) was prepared by dissolving 10 mg BS3 in 350 µL of 25 mM Sodium Phosphate, pH 7.4 buffer[62]. Using a 20-fold excess approach (20:1 Cross-linker:Protein), BS3 cross-linker solution was

added to the protein sample so that the final cross-linker concentration was between 0.5 to 5 mM. The sample was allowed to react at room temperature for 2 hours on ice. The unreacted BS3 was quenched with 60 mM Tris buffer and was allowed to react for 15 minutes at room temperature. Proteins of interest were then identified using SDS denaturing gel[63].

3.4.3 EspB Oligomerization at High Concentrations

Based on our chemical cross-linking experimental observations and mass spectrometric studies, EspB292 proteins form complex assemblies only at high concentrations (>100 μM) (**Figure 3.2-3.4**). Therefore, the cleaved EspB protein that is secreted into the extracellular milieu is capable of forming higher order oligomeric states and therefore, forming large protein assemblies when it accumulates in the small periplasmic space. Since the truncation of C-terminus of EspB is necessary to regulate the secretion of virulence factors, the protein-protein tetramer network is probably competing with four-helix-bundles for translocation (through the mycomembrane) using the same machinery.

3.4.4 EspB292 Assembly Formation in Presence of Heavy Metal Ion

Macrophages employ evolutionarily conserved mechanisms such as microbial ingestion, followed by maturation of the phagocytic vacuole (phagosome) to kill bacteria. During maturation process, macrophages utilize different mechanism to kill microbes, including phagosomal acidification and, the selective manipulation of various ions in the phagosomal lumen.

Phagosomal acidification is mainly established by the action of the vacuolar H^+-ATPase (v-ATPase) and the complementary roles played by other transporters and channels to counterbalance the ionic gradient of H^+ across the membrane (**Figure 3.5**). For example the influx of anions such as Cl^- via either the CFTR or the ClC family channels is the major

counterion transport that maintains vacuolar acidification. A counterflux of cations, such as the efflux of K^+, may also assist vacuolar acidification[64]. The immature mycobacterial phagosome in resting cells is considered to be mildly acidic (pH 6.5) but upon activation by the adaptive immune system, the pH is further decreased. Several lipids and proteins interfere with the phagosome maturation and resist the acidification[65]. To test whether EspB accumulation in periplasmic space would contribute to or be affected by the acidification machinery of the phagosomal milieu, the EspB aggregation was studied in different pHs ranging from 5 to 9. Our study reveals that pH variations don't affect the oligomeric state of EspB[56, 66].

Alternatively, the selective manipulation of various heavy metal ions in the phagocytic vacuole can lead to destroying the bacteria. Hereby, we will discuss the role of essential phagosomal nutrients and trace elements on the oligomeric state of the EspB292 protein.

Both the salts and ions can play a dominant role on the proton and electron gradients' maintenance that accompany phagosome acidification. For instance, several recent studies have indicated that within the first hour of phagocytosis of M. tuberculosis a decrease in vacuolar concentration of Ca^{2+} and K^+ can be observed which perhaps assists vacuolar acidification [13]. Within the next 24 hours the concentrations of these ions increase along with phagosomal concentration of sulphur and chlorine. Additionally, salts, metals and trace elements can impact the antimicrobial mechanisms and the biology of the pathogens. For example the movement of ions modifies cell activation and impacts the antimicrobial mechanism by inflammatory cytokines such as TNF and IFNγ.

3.5 Iron and TB

One of the metals that significantly affect the host–pathogen interactions of pathogenic intracellular microbes, such as Mycobacteria, Salmonella, and Legionella, is iron[5, 67]. Accumulation of the iron/manganese transporter at the phagosomal membrane leads to depletion of these metals at the phagosomal lumen within the first hours of infection. Iron depletion, thereby, arrests microbial intracellular development and constrains the microbial growth. In mycobacterium tuberculosis' phagosomal lumen, iron plays a biphasic behavior [68]. Its content first decreases up to four-fold within the first hours of infection, and increases from 5 to 25 folds during the following 24 hour[68-71]. Since high level of iron promotes the growth and infection of the mycobacterium tuberculosis, the iron chelation therapy has been recently studied to reduce the replication of Mtb and restore the host defense mechanism (**Figure 3.6**).

3.6 Zinc and Copper and TB

Other heavy metals that accumulate in large amounts inside the mycobacterial vacuole within the first 24 hours are Zn^{2+} and Cu^{2+}. Although several recent studies have revealed the role of phagosomal ions such as iron, zinc[69, 70, 72] and, copper[69, 73] and other chemicals in the control of pathogens, the information on the evolved mechanisms by which intracellular pathogens circumvent or exploit these conditions of host cell is limited. For instance, we know that to resist metal intoxication, microbes commonly up-regulate expression of efflux pumps of the P-type ATPase, but the role of these elements in phagosome biology remains elusive. One of the several questions that remain to be answered is: why Mtb deficient in expression of these transporters exhibits impaired intracellular survival? In other words, how does the content of any of these metal ions intoxicate or enhance the activity of the bacteria (**Figure 3.6**).

3.7 EspB Oligomerization in Presence of Iron and Copper

One key step to answer this question is to set up experiments that would mimic the phagosomal environment and study how the virulence factors are affected at each condition. To determine the effect of metal ions concentration on the EspB292 oligomerization, we compared the aggregation level of protein in various amounts of different metals. EspB exists mainly in its monomeric form in presence of many ions including Na^{+}, K^{+}, Cl^{-}, Ca^{2+}, Mg^{2+}, Po_4^{3-}, So_4^{2-} and Zn^{2+}. However, our findings further suggest that the aggregation of EspB292 can be significantly influenced by the content of the iron, and copper ions in the environment (**Figure 3.7-3.10**). Since the aggregation of EspB292 can alter the controlled secretion of essential virulence factors such as CFP-10 and ESAT-6, the maintenance or increase of the heavy metal concentration can ensure Mtb's survival or promotion.

3.8 Summary

Our findings suggest that the aggregation of EspB292 to tetramers neither depends on the pH of the periplasmic space nor is affected by ionic concentrations of Na^{+}, K^{+}, Cl^{-}, Ca^{2+}, Mg^{2+}, Po_4^{3-}, So_4^{2-} and Zn^{2+}. Nevertheless high level of heavy metal content of phagosomal environment within 24 hours of the infection can highly influence the aggregation of the EspB to higher order oligomeric complexes. Additionally, chemical cross-linking, and mass spectrometry studies confirm the formation of protein complex assemblies at high concentrations. Our analysis suggests functional roles for EspB provides insight into the pair-wise dependency of the EspB/Rv3879c secretion and its transmembrane translocation.

This study investigated the crystal structure and potential functions of the early secretory protein B in the Mtb progression. Our findings suggest that this protein has 16 monomers in the

asymmetric unit that is arranged into a set of four homo-tetramers and therefore it forms high order tetrameric aggregations upon cleavage at the periplasmic space of the mycobacterium by mycP1. Since the cleavage of EspB is essential for the regulation of virulence factor's secretion, we studied the phagosomal environmental conditions to investigate the elements that would impact the oligomerization of EspB.

Our findings suggest that the aggregation of EspB292 to tetramers depends on the protein's concentration. Therefore, upon EspB translocation to the small periplasmic milieu it would stay as a monomer. Upon accumulation to a high enough concentration the truncation of C-terminus can induce the protein-protein tetramer network formation which may further alter outcome of *M. tuberculosis* macrophage infection.

With regards to the secretory proteins of the ESX-1 secretion pathway, the crystal structure of EspB also indicates that the secretory protein pairs have a preference to adopt a four-helix-bundle structure. While the CFP-10/ESAT-6 and EspA/EspC pair was previously reported to have such a structure, the present study and previous in silico analysis by mande et al. suggests that the other secretory protein pairs (EspB/Rv3879c) also adopt similar structures.

We also observed that rather than C-terminus truncation of EspB by MycP1 protease, other conditions such as high heavy metal environment can also trigger the proteins aggregation and affect the outcome of the host-pathogen interactions to promote the disease. In recent years several studies have revealed the role of phagosomal ions such as iron, manganese, zinc and, copper and other chemicals in the control of pathogens. Unfortunately, the information on the evolved mechanisms by which intracellular pathogens circumvent or exploit these conditions of host cell is limited. Such mechanistic information can provide useful insights to develop better antimicrobial therapies in the future. For instance, we know that to resist metal intoxication,

microbes commonly up-regulate expression of efflux pumps of the P-type ATPase, but the role of these elements in phagosome biology remains elusive.

One other crucial aspects of the ESX-1 secretion mechanism pertain to the translocation of secretory proteins through the mycomembrane. In spite of considerable experimental efforts to understand the ESX-1 secretion system, the components involved in protein translocation through the mycomembrane is currently not known. In conclusion, a better knowledge of how ancient mechanisms involved evolved to perform host defense and microbial virulence functions will undoubtedly help understand host–pathogen interactions.

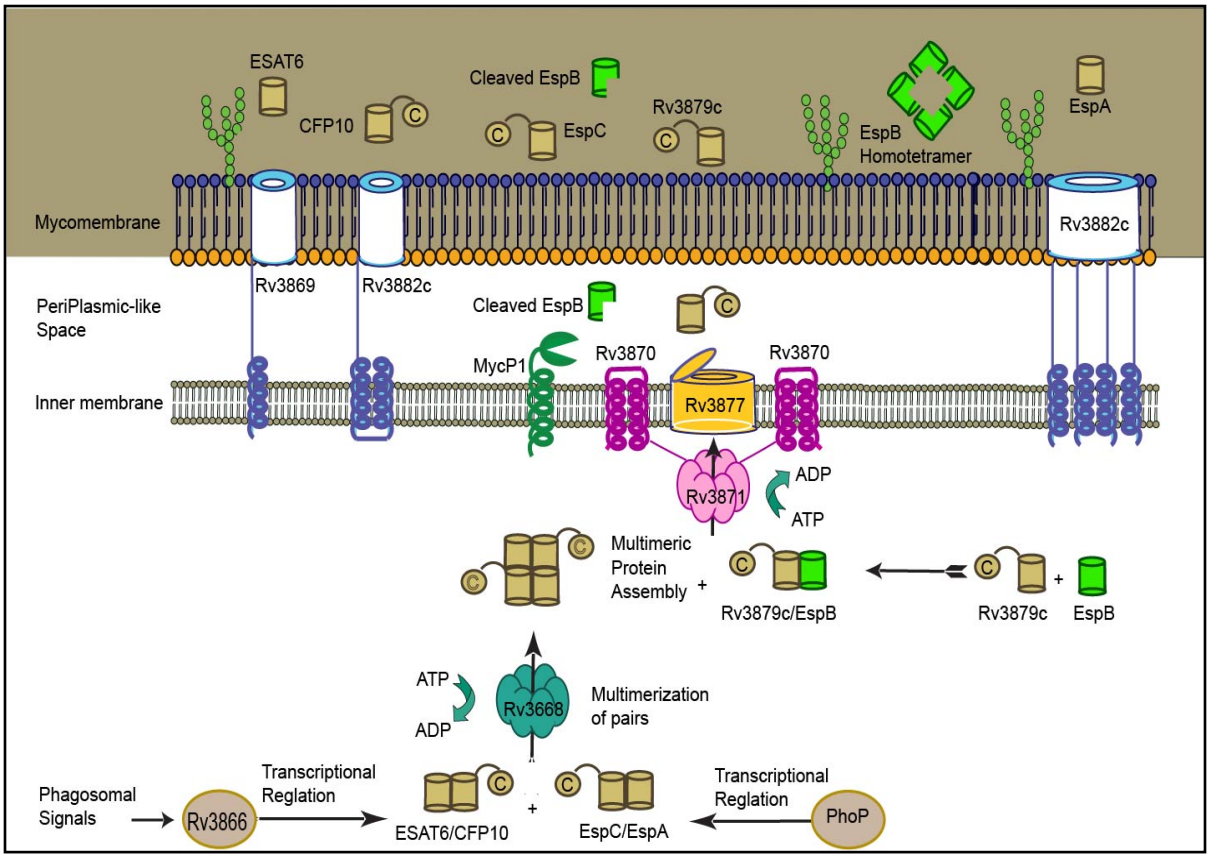

Figure 3.1 Proposed model of T7SS that is encoded by ESX-1: The phagosomal signals initialize the transcription and secretion of two pairs EspA/EspC and CFP-10/ESAT-6. Once expressed, Rv3868 assembles these pairs into multimers that are subsequently recognized by Rv3871 and Rv3870. These two FtsK family proteins form a translocase pore that the multimeric protein assembly is likely to pass through. Thereafter, the interaction of the N-terminal gate-like domain of Rv3877 with MycP1, a serine protease, opens the pore, thus enabling translocation of the substrate assembly from cytoplasm to the periplasm-like space. Ultimately, the substrate assembly crosses the mycomembrane through a channel formed by Rv3869, Rv3882c and the four proteins of MCE cluster 1 (Mce1B, Mce1C, Mce1F and Rv0177).

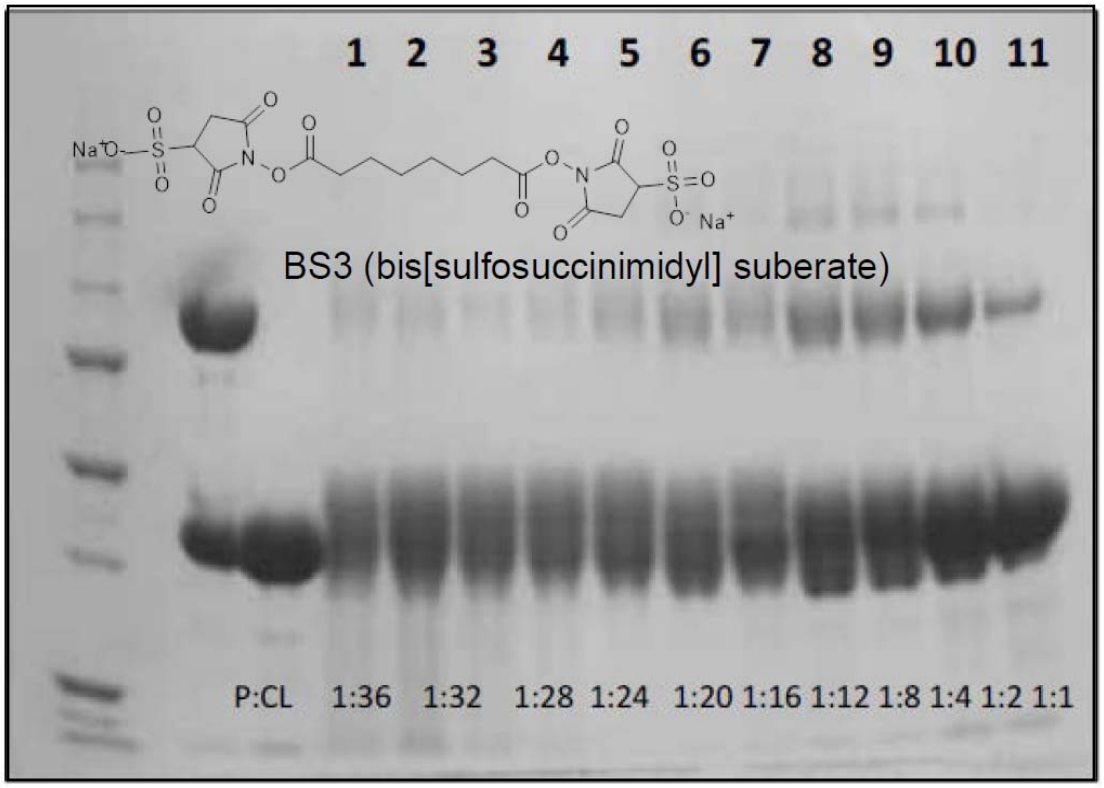

Figure 3.2 EspB292 SDS gel in presence of cross-linker: BS3 with respective spacer arm length of 11.4Å was used as the cross-linking agent that could covalently link two interacting proteins. Lanes 3 to 13 display the formation of protein assemblies as we decreased the crosslinker concentration (from 1.92 mM to 53μM). To all lanes except lane 1 BME was added as the reducing agent. The protein concentration was kept constant at 53μM in all lanes.

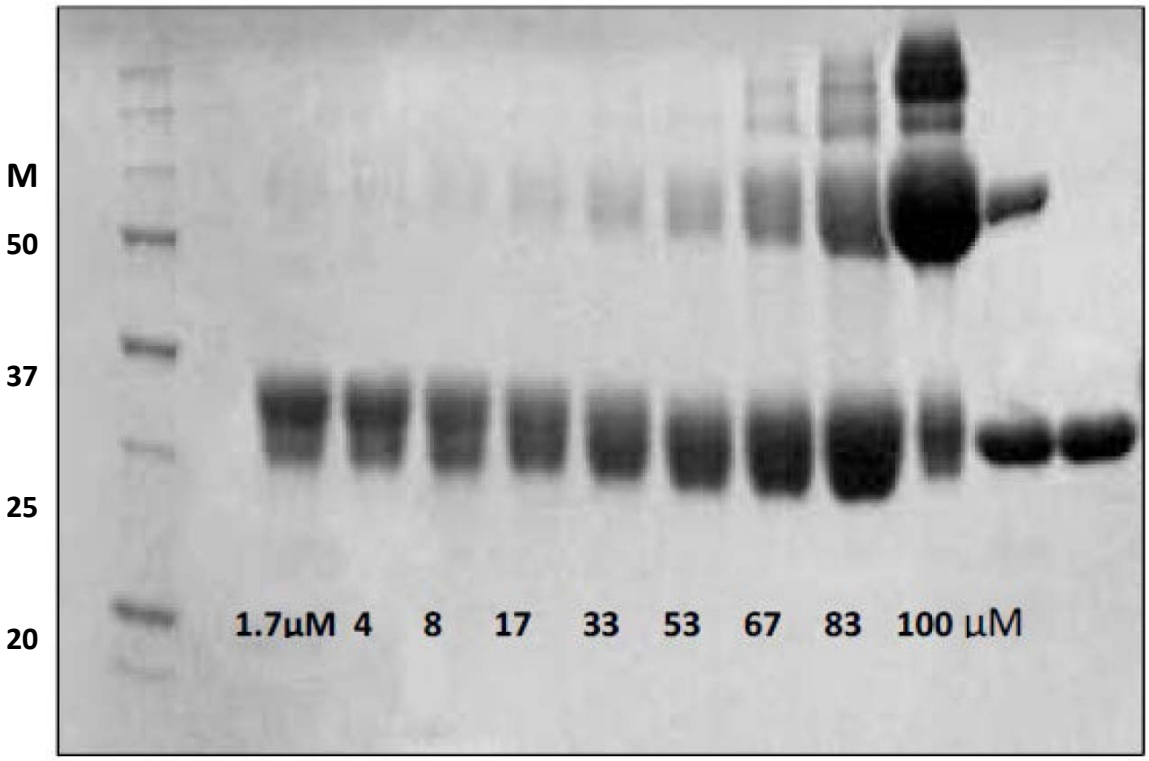

Figure 3.3 EspB292 SDS gel in presence of cross-linker: Lanes 1 to 9 displays the formation of protein assemblies as we increased the protein concentration (1µM-100µM). To all lanes except lane 10 BME was added as the reducing agent. Lanes 10 and 11 are showing the pure protein without the crosslinker. Numbers between panels indicate sizes in kilodaltons. The protein:Crosslinker ratio was kept constant at in all lanes.

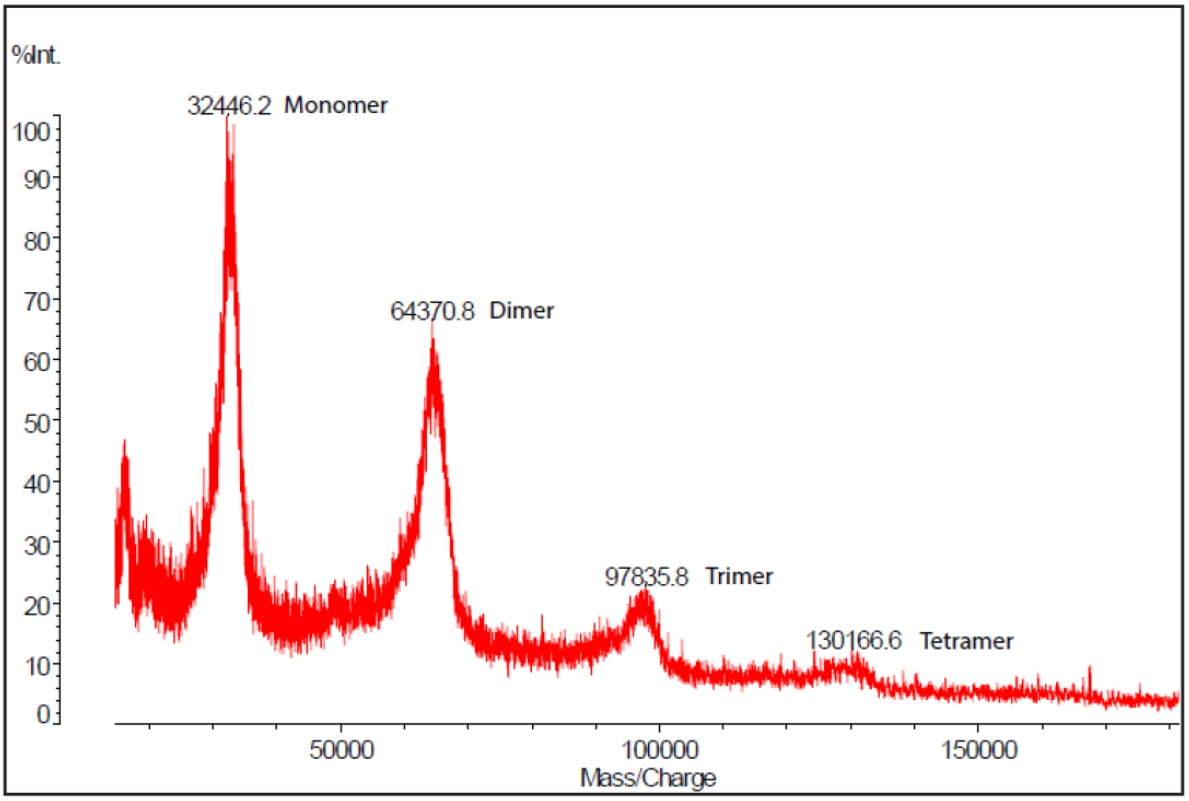

Figure 3.4 Mass spectrometry of EspB292 proteins: The MS displays the formation of higher order complex assemblies at 50 µM.

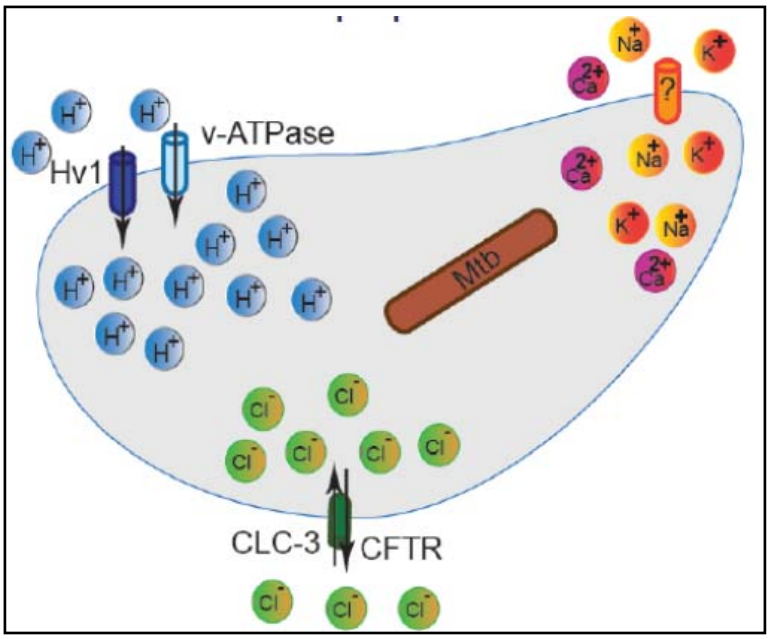

Figure 3.5 Phagosomal acidification: The acidification of phagosome is established by the action of the vacuolar H^+-ATPase (v-ATPase) and the complementary roles played by other transporters and channels to counterbalance the ionic gradient of H^+ across the membrane. For example the influx of anions such as Cl^- via either the CFTR or the ClC family channels is the major counter ion transport that maintains vacuolar acidification[74].

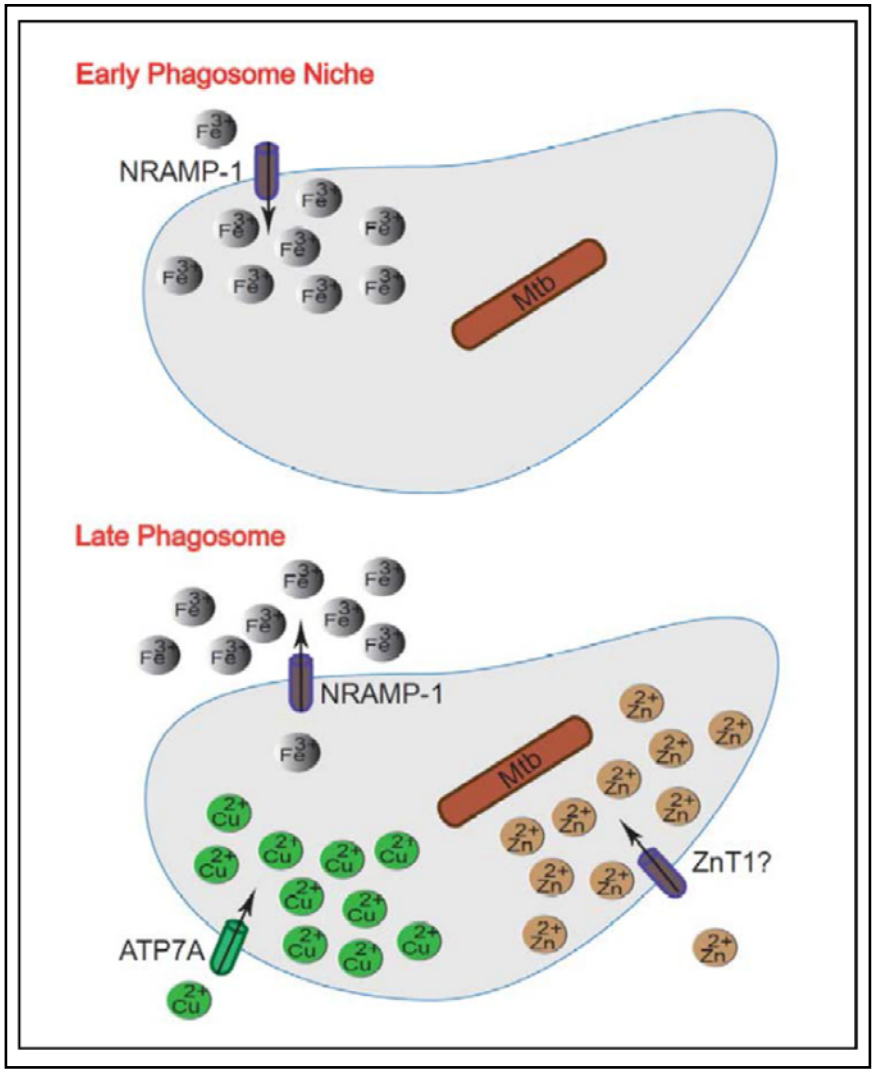

Figure 3.6 Iron, zinc, and copper concentration: Early phagosome contains high concentration of Fe^{3+} which decreases within the next 24 hours after phagocytosis. The decrease in iron III ion concentration happens along with the increase in the concentrations of Cu^{2+}, and Zn^{2+} ions as well as Sulphur and chlorine. Additionally, salts, metals and trace elements can impact the antimicrobial mechanisms and the biology of the pathogens.

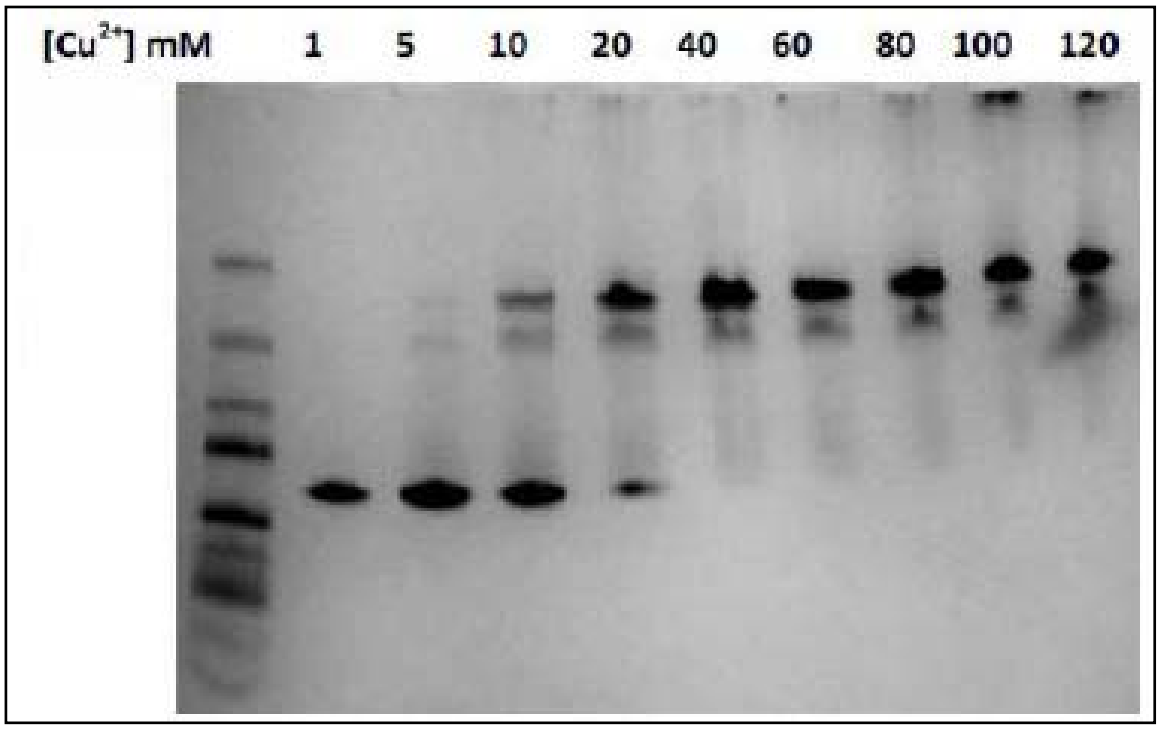

Figure 3.7 EspB292 assembles to higher order aggregation complexes upon increasing the concentration of copper from 1mM to 120mM: The native gel of 20μM EspB at all lanes in presence of increasing concentrations of $CuCl_2$.

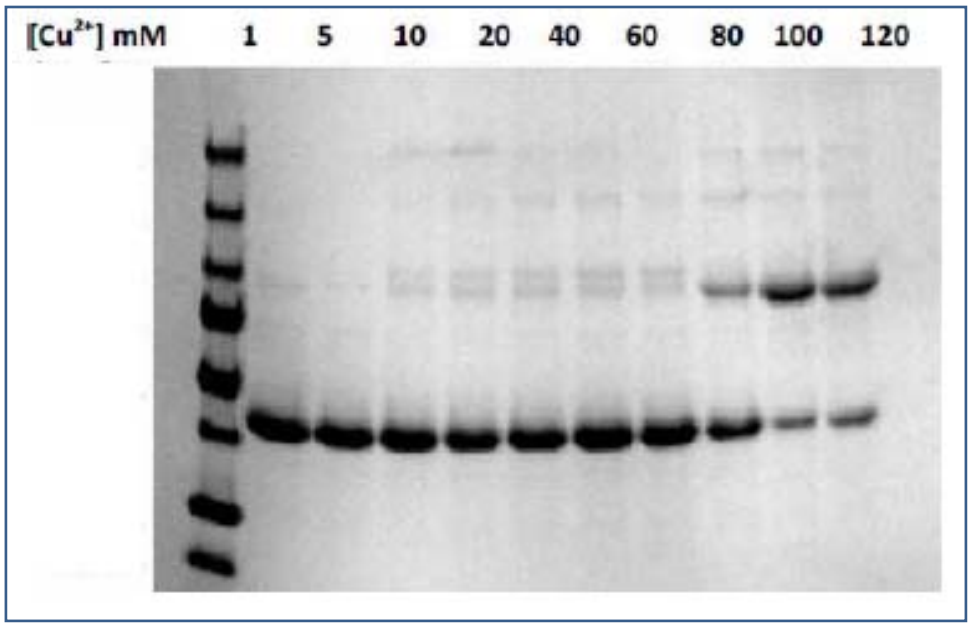

Figure 3.8 Denaturing gel of 20μM EspB and 2.5mM BS3 cross-linker at all lanes in presence of increasing concentrations of $CuCl_2$.

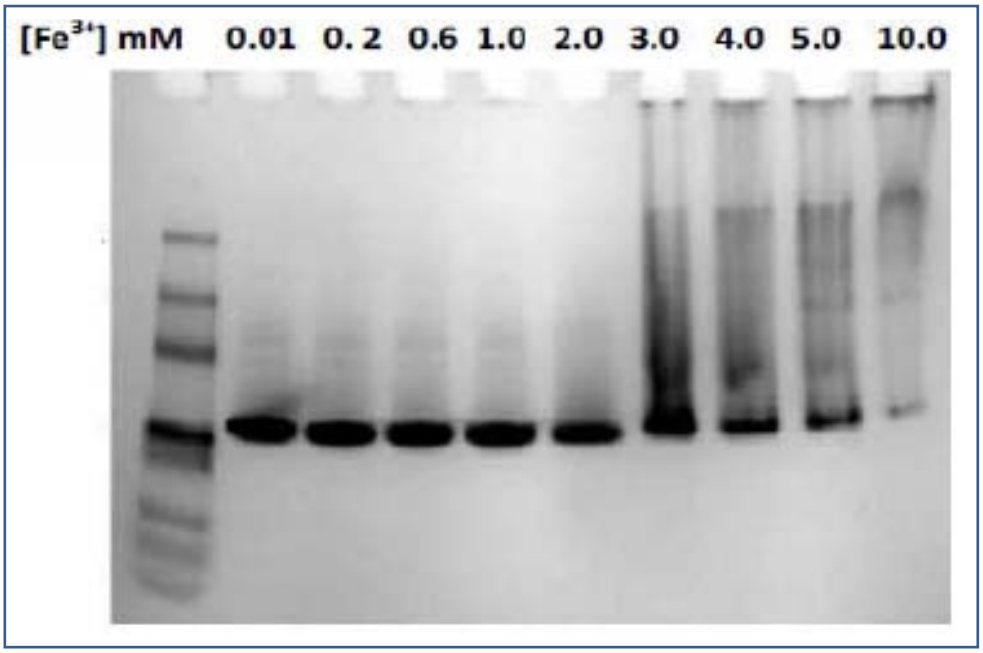

Figure 3.9 EspB292 assembles to higher order aggregation complexes upon increasing the concentration of iron (III) from 10µM to 10mM: The native gel of 20µM EspB at all lanes in presence of increasing concentrations of $FeCl_3$.

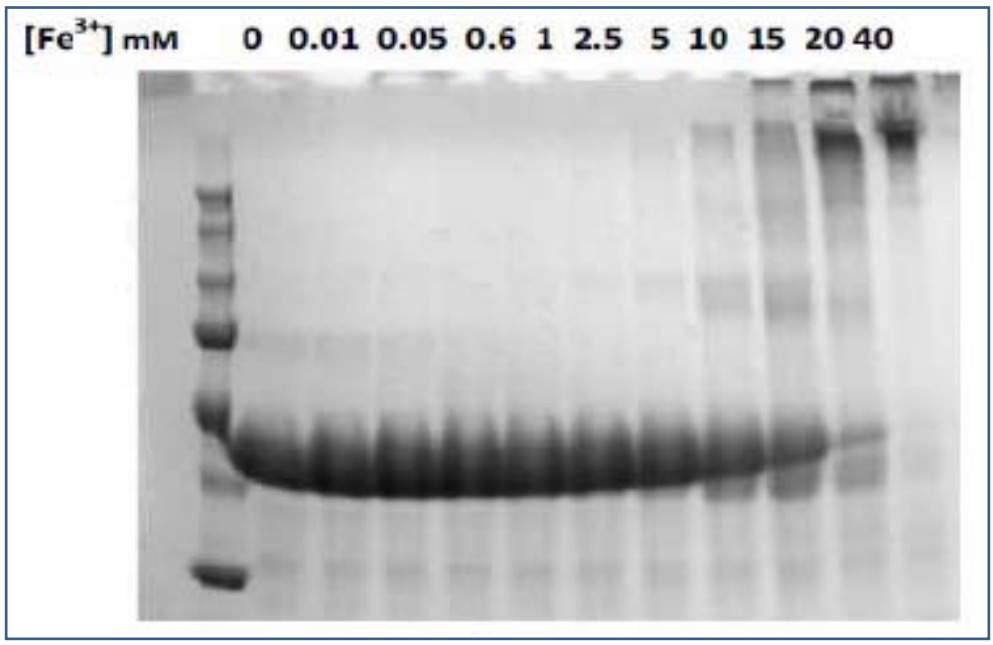

Figure 3.10 Denaturing gel of 20μM EspB and 2.5mM BS3 cross-linker at all lanes in presence of increasing concentrations of $FeCl_3$.

Chapter 4: Purification and Crystallization of Ct-Rio2

The purification of Rio2 family of proteins in eukaryotes has been tried by my colleague Vatsala Sagar for Rio2p and hRio2 and the efforts to obtain the structure of a eukaryotic Rio2 failed[75]. In the case of Rio2p the crystals were small, and in the case of hRio2 poor solubility prevented crystal screening[76].

4.1 Expression of the Ct-Rio Protein

Ever since the identification of Rio proteins in yeast, several Rio proteins have been expressed and purified in *Escherichia coli*, and their enzymatic properties and structural features were characterized. For instance, *A. fulgidus* Rio1 was expressed in *E. coli* as a fusion protein containing the N-terminal His tag followed by a tobacco etch virus (TEV) protease cleavage site. The tag was subsequently removed during protein purification and the enzyme is shown to be capable of autophosphorylation on Ser-108. *A. fulgidus* Rio2 with its native N terminus was also expressed and purified[77], and it was also shown to be active in an autophosphorylation assay, although at a level much lower than that of Rio1. The site of autophosphorylation was determined to be Ser-128[78].

In this chapter, I discuss the purification and crystallization of Rio2 from a thermophilic fungus, *Chaetomium thermophilum* (Ct-Rio2). The construct shown in **Figure 4.1**, was obtained from the laboratory of Dr. Ed Hurt in the Biochemistry Center, Heidelberg University, Germany. There were issues in the purification and crystallization of Ct-Rio2: it co-purified with a close molecular weight degradation product; it nonspecifically bound to RNA in the lysate; and crystals of Ct-Rio2 were small and did not diffract well. These problems were corrected by overloading the His Trap™ column, which caused Ct-Rio2 to elute differently from the close

molecular weight contaminant, using a Q™ column to separate RNA from ctRio2, and seeding with microcrystals to produce larger, better diffracting crystals (for more information the reader is referred to my colleague Vatsala Sagar's dissertation).

4.2 Purification of Rio2 from *Chaetomium Thermophilum*

Following the optimized procedure outlined below Ct-Rio2 was designed and was further expressed, purified and crystallized. The p7-Ct-Rio2 vector (from the Hurt lab), transformed into *E. coli* Rosetta™(DE3) pLysS (Novagen), and was used to inoculate 100 mL culture which was expressed with 34 μg/mL chloramphenicol and 100 μg/mL ampicillin. The cells were grown at 37°C while shaking at 250 rpm. The overnight culture was diluted 1/100 into four 4 L baffle flasks containing 1 L of LB with ampicillin and chloramphenicol. The four flasks were then incubated at 37°C with shaking at 250 rpm. Once the OD_{600} reached 0.6-0.8, 1 mM of Isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to the media for induction and expression proceeded at 18°C overnight. The cells were centrifuged at 5,000 rpm and 4°C for 20 minutes using a Sorvall centrifuge (Dupont) with a SLC 4000 rotor. The bacterial pellet was stored at -80°C.

The cells were centrifuged at 5,000 rpm for thirty minutes. The pellet was about 4-5 grams per liter of cell culture. The cell pellets were frozen at -80°C. The frozen pellet was resuspended in 10 mL of Buffer A (200 mM NaCl , 10% glycerol, 50 mM Tris, pH 8.0, 2.5 mM $MgCl_2$, 0.05% BME) per gram of cell pellet. 5 mg of DNAse, one Complete, EDTA free, protease inhibitor cocktail tablet (Roche), 2 mg of RNase A, 1ml of 10X Bugbuster™ and 0.5 mg of lysozyme/mL per liter of cell pellet were added to the solution. The cells were stirred at 4°C for thirty minutes. The lysate was, subsequently, centrifuged at 16,000 rpm for fifty minutes

to pellet the cell debris. The supernatant containing the soluble protein was filtered through a 0.2 micron filter.

The supernatant was passed over a GE Healthcare 5 mL His Trap™ column that was pre-equilibrated with Buffer A. Only one His Trap™ column was used in order to overload the column with the supernatant. Overloading the column eluted the Ct-Rio2 at 30 mM imidazole and in pure and concentrated fractions. Overloading the column with protein allows the elimination of protein contaminants including the degradation product. Since Ct-Rio2 has higher affinity for the nickel column and binding sites on the nickel column are limited, Ct-Rio2 will compete with degradation products for binding to the column. The result is that the major protein contaminants will not bind to the column. Buffer B (200 mM NaCl, 10% glycerol, 50 mM Tris, pH 8.0, 2.5 mM $MgCl_2$, 0.5 M imidazole, 0.05% BME) was used to elute the protein. 6% Buffer B was used to elute CtRio2 from the column. The column was washed with 20% Buffer B and then 100% Buffer B (**Figure 4.2**).

The concentrated and pure fractions were pooled together for TEV cleavage. Ct-Rio2 was concentrated to 10 mg/mL on a millipore centrifugal filter unit at 4°C. 10 mg of TEV protease was added to the Ct-Rio2, which was then dialyzed overnight in Buffer A at 4°C. In order to cleave the His tag from the N-terminus for each 1 mg/ml of concentration of Ct-Rio2, 1 mg/ml of TEV protease was necessary. With these concentrations most of Ct-Rio2 was found to be cleaved and eluted in the flow-through.

The next day, the His Trap™ column was equilibrated with Buffer A. Ct-Rio2 was passed over the column and the flow-through was collected and assumed to be cleaved protein. The Ct-Rio2 was concentrated to 10 mg/mL as determined by A280.

4.3 Ct-Rio2 Binds to RNA Nonspecifically

The Ct-Rio2 protein tends to binds to RNA and this binding makes purification steps a little difficult. There are two solutions for purifying the protein and get rid of RNA contaminants [48]. One is to add 2 mg of RNase A to the lysate in the lysis buffer and second is to pass the protein over an anion exchange Q^{TM} after the second His trap column during the last step of purification of Ct-Rio2. In the chromatogram of the elution of the protein from the Q^{TM} column two peaks appear, one for ctRio2 at a lower concentration of 300 mM salt, and one for RNA at a higher concentration of 500 mM salt. The second peak is RNA because it has a high affinity for the Q^{TM} column. Its A_{260}/A_{280} absorbance ratio is over 2 and it shows no protein bands on the gel.

The Q^{TM} column is an important step for purification. If RNA continues to remain bound to ctRio2, it would prevent Ct-Rio2 from forming crystals. The RNA is bound nonspecifically and consists of a variety of different RNAs. Any heterogeneity in the protein sample precludes the protein from crystallization. The pure protein was concentrated to 10mg/ml and was kept in -80ºC.

The cleaved protein was further purified by size exclusion chromatography one day before setting up the crystal screening trays using a Superdex 200 column (GE healthcare). The column was equilibrated in 50 mM Tris pH 8, 200 mM NaCl and 0.05% β-mercaptoethanol (**Figure 4.3**). All purification steps were monitored by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS PAGE) for purity.

4.4 Crystallization of Ct-Rio2 in Presence of ADP

Purified Ct-Rio2 was concentrated to 10 mg/mL and was subjected to robotic sparse matrix screening using Phoenix Liquid Handing System (Art Robbins). Commercially available

matrix screens were used including Qiagen's PEG Suite, Hampton's Index Screen, Emerald Biosystems Wizard I, II, and III Screens, Hampton Research's Natrix Screen and Qiagen's Cryos Suite. Initial crystal screening was performed in 96 well plates with three different proteins to well solution ratios (1:2, 1:1, 2:1). The protein was co-crystallized with ADP/BeF_2 and ADP/Na_2VO_3. The structure was expected to mimic the transition state of the phosphoryl transfer from ATP to Asp257 in Rio2's active site and the subsequent hydrolysis of the aspartyl phosphate that could power late cytoplasmic 40S subunit biogenesis.

4.5 Crystallization of Ct-Rio2 in Presence of ADP/$BeCl_2$/KF

Ct-Rio2 was crystallized in the presence of ADP, $MgCl_2$, $BeCl_2$ and KF. After setting up the initial crystal screenings the crystals started to show up in two weeks. Several crystals were obtained for each of the complexes. We obtained the diffraction data sets for Ct-Rio_2/ADP/BeF_2 at 4 conditions with resolutions of 2.16-3.31Å (**Figure 4.4**). The best crystal belonged to the Index screen condition number 39 (0.1 M HEPES pH 7.0, 30% v/v Jeffamine ED-2001 pH 7.0).

Data was collected at 100K at the NE-CAT beamline at the Advanced Photon Source (APS), Argonne, IL, USA. The crystals diffracted to 2.381 Å. Data was indexed, integrated and scaled using iMosflm package program. The crystals belonged to R3: H space group and the unit cell dimensions were a=117.359 Å, b=117.359 Å, c=70.898 Å and α=90.000, β=90.000, γ=120.000 (**Table 4.1**). The overall multiplicity was 6.4 with 99.66% data completion. The structure was solved using an automated experimental phasing program, that is a part of PHENIX. Experimental phases were calculated using Phaser and improved by density modification using RESOLVE. After obtaining density modified phases the program carries out a round of model building and several rounds of molecular replacement and refinement, model

building were carried out in COOT. Waters were added to the structure using PHENIX and the model has a Rwork/Rfree of 18.39/24.9 (**Table 4.2-4.3**).

Crystal structures of the AMP bound complex of Ct-Rio2 were obtained. However, no ADP was observed in the active site of the enzyme. Surprisingly, the Ct-Rio2 was bound to AMP in the active site. This could be due to the hydrolysis of the ADP to AMP before the crystals were formed. We thought this issue could have occurred because of the pH of the solution. Testing the pH of the original solutions that were used to set up crystal screening trays showed that the pH was as low as 5.4.

In the next procedure, 1M Hepes buffer at pH of 7 was used to set the pH of the original ADP solution to 7. Subsequently, the newly prepared ADP was incubated with Ct-Rio2 for one hour in the ice. This protein solution was then, subjected to robotic sparse matrix screening using Phoenix Liquid Handing System again. Commercially available matrix screens were used including Qiagen's PEG Suite, Hampton's Index Screen, Emerald Biosystems Wizard I, II, and III Screens, Hampton Research's Natrix Screen and Qiagen's Cryos Suite. Similarly, initial crystal screening was performed in 96 well plates with three different proteins to well solution ratios (1:2, 1:1, 2:1). The protein was co-crystallized with ADP at pH 7, $BeCl_2$ and KF. The crystals that we obtained in the new procedure were of less resolution of 2.67 Å (**Figure 4.5**).

4.5.1 Data Collection and Structure Determination

Data was collected at 100K at the NE-CAT beamline at the Advanced Photon Source (APS), Argonne, IL, USA. Data was indexed, integrated and scaled using HKL2000 package program. The crystals with the AMP bound in the active site belonged to the R3 space group and the unit cell dimensions were a=119.658 Å, b=119.658 Å, c=72.342 Å and α=90.000, β=90.000, γ=120.000 (**Table 4.4**). The structure was solved using PHENIX. After several rounds of molecular replacement and refinement in PHENIX, waters were added to the structure. . The

overall multiplicity was 2.9 with 98.10% data completion. The final model had a Rwork/Rfree of 19.23/27.24 (**Table 4.5-4.6**).

Although crystals were grown with Ct-Rio2 in presence of ADP and $MgCl_2$ and $BeCl_2$/KF again the ADP couldn't be observed and the refined crystal structure didn't show any trace of $BeCl_2$/KF in the active site. Instead the ADP was cleaved to AMP and the protein was again bound to AMP in the structure (**Figure 4.6**).

4.6 Crystallization of Ct-Rio2 in Presence of ADP/Na_2VO_4

For the Ct-Rio2/ADP/Vanadate complex, we obtained the diffraction data sets at 7 conditions with resolutions of 2.02-3.0Å (**Figure 4.7**). The best crystal belonged to the Natrix screen condition number 43 (0.05 M Ammonium acetate, 0.01 M Magnesium chloride hexahydrate,0.05 M TRIS hydrochloride pH 7.5, 10% v/v (+/-)-2-Methyl-2,4-pentanediol) at 2.07 Å resolution.

4.6.1 Data collection and Structure Determination

Data was indexed, integrated and scaled using HKL2000 package program. The crystals belonged to R3 space group and the unit cell dimensions were a= 118.402 Å, b= 118.402 Å, c= 72.43 Å and α=90.000, β=90.000, γ=120.000 (**Table 4.7**). The structure was solved using PHENIX. The initial model was first subjected to two rounds of molecular replacement using phaser that is part of PHENIX. Subsequent rounds of model building were carried out in COOT accompanied by several rounds of refinement using phenix-refine. Waters were added to the structure using PHENIX. The overall multiplicity was 3.6 with 99.93% data completion. The final model had a Rwork/Rfree of 20.17/25.11 (**Table 4.8**).

Although crystals were grown with CT-Rio2 in presence of ADP, $MgCl_2$ and sodium orthovanadate, the refined crystal structure didn't show any trace of ADP in the active site.

Instead the ADP was again cleaved to AMP and the protein was bound to AMP in the structure. Again, contact was observed between the AMP α-phosphate and Asp257and Mg^{2+} (**Figure 4.8**).

4.6.2 Crystallization of Ct-Rio2 in Presence of ADP/VO_4 at pH 7

These findings were odd and therefore, we thought the cleavage could be due to the acidic pH of the droplets. Nucleotides such as ATP and ADP are unstable at lower pH and tend to release inorganic phosphate. Therefore, new crystal screenings were performed while the pH of the solutions was monitored to be at 7. The crystals were also grown in a hanging drop screen to make larger crystals. Crystals didn't grow in the new buffer when pH was set to 7 and the only crystal that was formed grew large enough for X-ray diffraction studies after 3 weeks at 20°C. The crystal was sent for data collection and it diffracted to 3.25 Å (**Figure 4.9**).

4.6.3 Data collection and structure determination

Data was indexed, integrated and scaled using HKL2000 package program. The crystals belonged to R3 space group and the unit cell dimensions were a= 119.252 Å, b= 119.252 Å, c= 72.75 Å and α=90.000, β=90.000, γ=120.000 (**Table 4.9**). The structure was solved using PHENIX. Subsequent rounds of model building were carried out in COOT accompanied by several rounds of refinement using phenix-refine. Waters were added to the structure using PHENIX. The new crystals with CT-Rio2 in presence of ADP, $MgCl_2$ and sodium orthovanadate did show ADP in the active site.

The crystal structure of Ct-Rio2 showed a bilobal kinase domain (green and yellow) connected to a N-terminal winged helix domain (pink). The "hinge" region connects the N- and C- terminal lobes of the kinase domain. The ADP binding loop "P-loop", metal binding loop used to bind Mg, and catalytic loop which contains the catalytic D232 are indicated by arrows (**Figure 4.10**).

The overlay of the C-Terminal domain in the crystal structures of Ct-Rio2 bound to sodium orthovanadate showed that the winged region was not superimposed perfectly and the presence of ADP opened up the phosphate binding loop. Crystal structure of Ct-Rio2 at pH 7 shows the vanadate and ADP in the active site bound to the Asp257 and Mg^{2+}. The aspartate257 seems to be binding to both VO_4 and Mg^{2+} through H-bonding interactions. In the bilobal kinase domain of Ct-Rio2 at pH 5.4, AMP is bound with no trace of VO_4 in its active site. The winged helix domain of the two structures is not completely superimposed. The overlay of the AMP region in the crystal structures of Ct-Rio2 bound to AMP and ADP at two different pH showed that both C and N terminal lobes were apart. The overlay of the active site of Ct-Rio2 bound to ADP and Orthovanadate vs. Ct-Rio2 bound to AMP displays a dramatic movement of Asp257 from the AMP bound complex to the vanadate bound complex.

4.7 Summary

Ct-Rio2 was expressed and purified on His Trap^{TM}, Q^{TM} column and size exclusion column. Success was achieved by overloading the His Trap^{TM} column to remove major protein contaminants and degradation products. TEV cleavage was optimized to cleave most of the Ct-Rio2 in solution by adding 1mg/ml TEv for each 1 mg/Ml od the protein. Passing of Ct-Rio2 over the His Trap^{TM} a second time after TEV cleavage, removed the remaining contaminants. Ct-Rio2 was next purified on a Q^{TM} column to remove non-specifically bound RNA that could impede its crystallization. Following this procedure, the cleaved protein was further purified by size exclusion chromatography one day before setting up the crystal screening trays using a Superdex 200 column (GE healthcare). Crystals of Ct-Rio2 were then obtained in presence of AMP and ADP/VO_3^{2-} at the synchrotron, and finally a structure solution was obtained.

Ct-Rio2 construct:

Figure 4.1 Schematic representation of Ct-Rio2 construct: Ct-Rio2 construct is designed with an N-terminal 6x His tag followed by a TEV cleavage site.

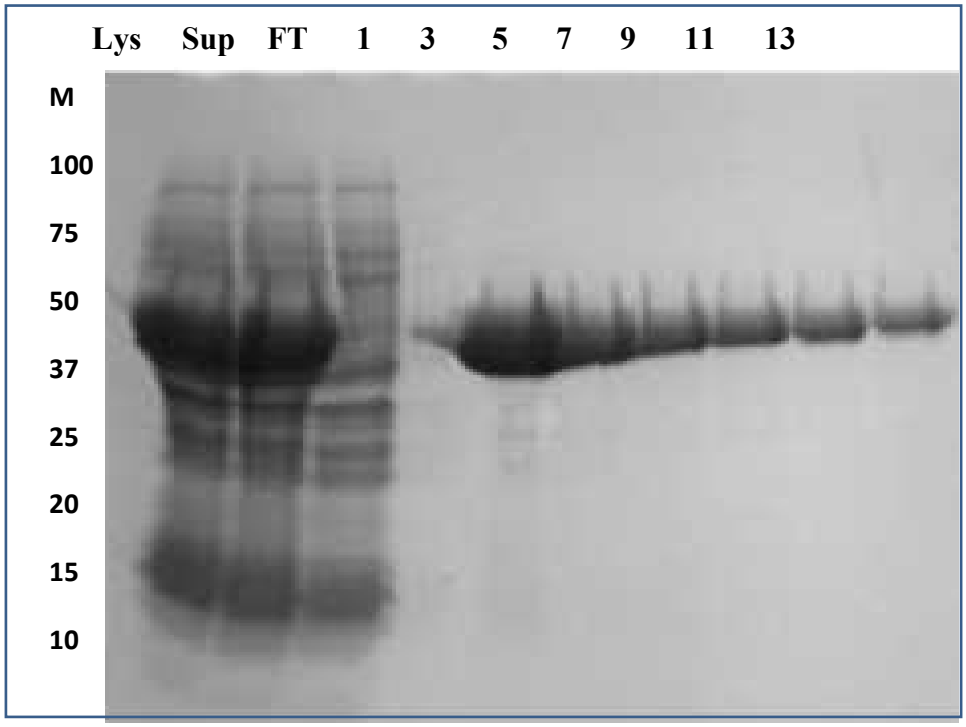

Figure 4.2 Purification of Ct-Rio2 on His Trap™ Column: Supernatant containing overexpressed Ct-Rio2 was passed over one His Trap™ column. The SDS-PAGE gels, stained with Coomassie Blue, is labeled with lysate (Lys), supernatant (Sup), and flow through (FT) the fraction numbers 1-7 that represent 30 mM imidazole concentration and 9-13 represent 50 mM imidazole at which the fractions eluted. Ct-Rio2 begins to elute from the His Trap™ column at 30 mM imidazole.

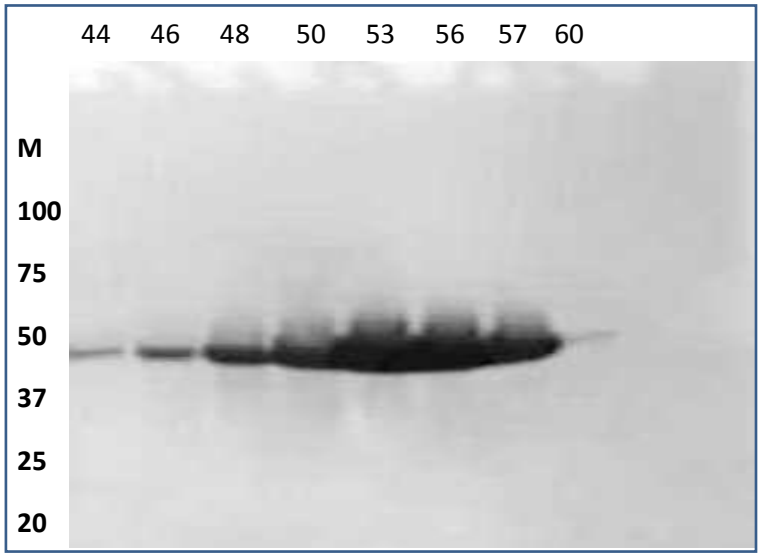

Figure 4.3 Preparative gels after size exclusion chromatography: Lane M is the protein ladder. Gel of Ct-Rio2 displays pure protein at the expected molecular weight of 44.6 KDa.

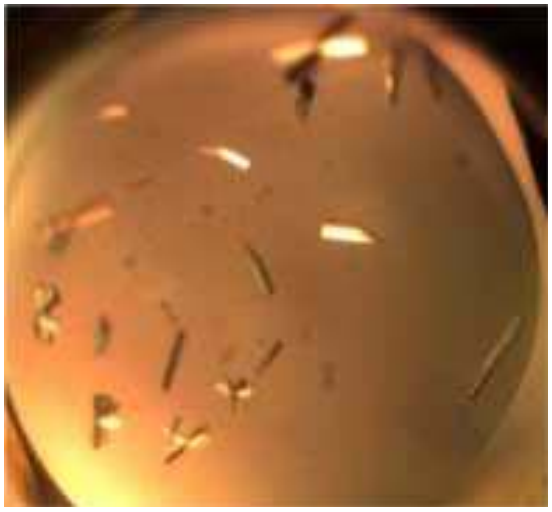

Figure 4.4 The Ct-Rio2 crystals formed in the presence of ADP/$BeCl_2$/KF in the hanging drop at pH 5.4: The crystals were produced in 0.1M HEPES pH 7.0., and 30% v/v jeffamine ED-2001 pH 7.0 at the 2:1 protein to well solution ratio.

Table 4.1 Data collection and current refinement statistics for AMP-bound Ct-Rio2 at the pH 5.4.

Resolution range (Å)	58.15 - 2.381 (2.466 - 2.381)
Space group	R 3: H
Unit cell	117.359 117.359 70.898 90 90 120
Total reflections	93647 (9349)
Unique reflections	14540 (1445)
Multiplicity	6.4 (6.5)
Completeness (%)	99.66 (99.31)
Mean I/sigma (I)	19.16 (1.75)
Wilson B-factor	47.84
R-merge	0.08204 (1.2)
R-meas	0.08926
CC1/2	0.999 (0.553)
CC*	1 (0.844)
R-work	0.1839 (0.2601)
R-free	0.2409 (0.3324)
Number of non-hydrogen atoms	2573
Macromolecules	2541
Ligands	32
Protein residues	343
RMS (bonds)	0.009
RMS (angles)	1.33
Ramachandran favored (%)	95
Ramachandran outliers (%)	2.1
Clash score	8.36
Average B-factor	66.40

Table 4.2 AMP bound Ct-Rio2 Scalepack Logfile at pH 5.4. The Ct-Rio2 crystals were formed in the presence of ADP/$BeCl_2$/KF in the hanging drop. The crystals were produced in 0.1M HEPES pH 7.0. 30% v/v jeffamine ED-2001 pH 7.0 at the 2:1 protein to well solution ratio.

	Overall	InnerShell	OuterShell
Low resolution limit	59.23	59.23	1.83
High resolution limit	1.80	9.88	1.80
Rmerge (within I+/I-)	1.122	0.116	84.243
Rmerge (all I+ and I-)	0.280	0.040	21.537
Rmeas (within I+/I-)	1.586	0.164	119.137
Rmeas (all I+ & I-)	0.396	0.057	30.458
Rpim (within I+/I-)	1.122	0.116	84.243
Rpim (all I+ & I-)	0.280	0.040	21.537
Rmerge in top intensity	0.073	-	-
Total number of observations	161653	1170	6709
Total number unique	91794	612	4021
Mean((I)/sd(I))	20.7	34.1	1.2
Mn(I) half-set correlation CC(1/2)	0.919	0.982	-0.030
Completeness	87.4	97.6	77.9
Multiplicity	1.8	1.9	1.7
Anomalous completeness	55.8	86.1	40.6
Anomalous multiplicity	0.6	0.9	0.5
DelAnom correlation between half-sets	0.000	0.000	0.000
Mid-Slope of Anom Normal Probability	1.478	-	-

Table 4.3 AMP-bound Ct-Rio2 phenix.xtriage logfile at pH 5.4. The Ct-Rio2 crystals were formed in the presence of ADP/$BeCl_2$/KF in the hanging drop. The crystals were produced in 0.1M HEPES pH 7.0. 30% v/v jeffamine ED-2001 pH 7.0 at the 2:1 protein to well solution ratio.

Res. Range	I/sigI>1	I/sigI>2	I/sigI>3	I/sigI>5	I/sigI>10	I/sigI>15
58.18 - 4.72	99.7%	99.4%	98.9%	97.5%	95.1%	92.1%
4.72 - 3.75	99.8%	98.5%	97.2%	94.8%	88.6%	80.6%
3.75 - 3.27	99.0%	95.5%	91.9%	84.2%	70.7%	56.7%
3.27 - 2.97	96.8%	89.1%	81.5%	67.4%	42.0%	26.6%
2.97 - 2.76	93.4%	76.0%	62.6%	42.9%	17.6%	8.4%
2.76 - 2.60	89.6%	62.9%	45.6%	24.5%	6.4%	2.0%
2.60 - 2.47	85.4%	48.0%	28.0%	12.1%	2.5%	0.9%
2.47 - 2.36	81.5%	38.1%	18.4%	6.5%	0.6%	0.8%
2.36 - 2.27	76.3%	29.1%	11.3%	2.4%	0.2%	0.0%
2.27 - 2.19	72.9%	20.6%	5.3%	1.1%	0.3%	0.0%
2.19 - 2.12	67.9%	14.1%	3.1%	0.3%	0.0%	0.0%
2.12 - 2.06	63.7%	10.5%	1.0%	0.0%	0.0%	0.0%
2.06 - 2.01	60.6%	8.5%	0.5%	0.0%	0.0%	0.0%
2.01 - 1.96	48.6%	5.2%	0.3%	0.0%	0.0%	0.0%

The completeness of data for which I/sig(I)>3.00, exceeds 85%
for resolution ranges lower than 3.27A.

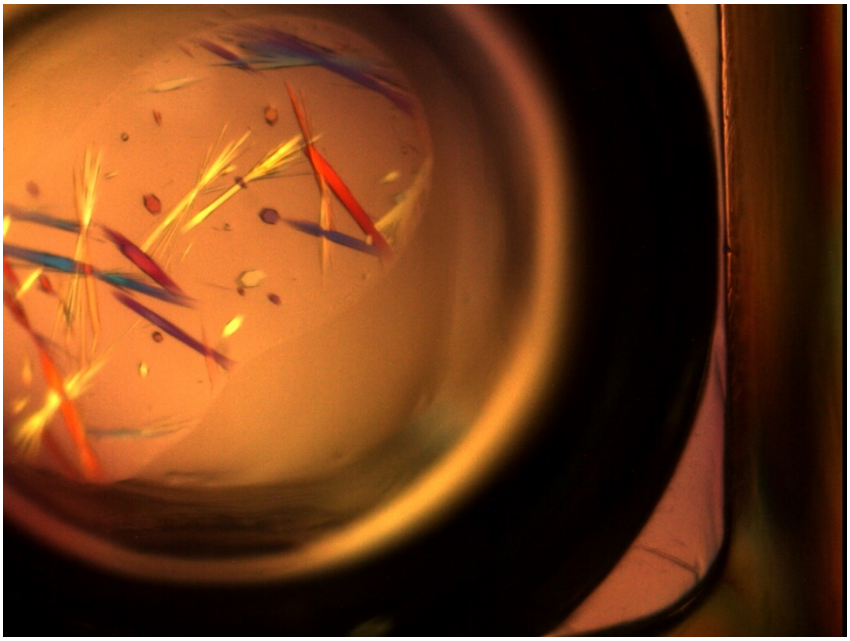

Figure 4.5 The CtRio2 crystals formed in the presence of ADP/$BeCl_2$/KF in the hanging drop at pH 7: The crystals were produced in 0.1M HEPES pH 7.0. 30% v/v jeffamine ED-2001 pH 7.0 at the 2:1 protein to well solution ratio.

Table 4.4 Data collection and current refinement statistics for AMP-bound Ct-Rio2 at the pH of 7. The Ct-Rio2 crystals were formed in the presence of ADP/$BeCl_2$/KF in the hanging drop. The crystals were produced in 0.1M HEPES pH 7.0. 30% v/v jeffamine ED-2001 pH 7.0 at the 2:1 protein to well solution ratio.

Resolution range (Å)	59.83 - 2.67 (2.766 - 2.67)
Space group	R 3: H
Unit cell	119.658 119.658 72.342 90 90 120
Total reflections	31595 (3181)
Unique reflections	10772 (1101)
Multiplicity	2.9 (2.9)
Completeness (%)	98.10 (97.83)
Mean I/sigma (I)	6.08 (0.39)
Wilson B-factor	69.38
R-merge	0.1973 (2.943)
R-meas	0.2448
CC1/2	0.98 (0.0356)
CC*	0.995 (0.262)
R-work	0.1923 (0.3031)
R-free	0.2724 (0.4227)
Number of non-hydrogen atoms	2573
Macromolecules	2541
Ligands	32
Protein residues	343
RMS (bonds)	0.010
RMS (angles)	1.42
Ramachandran favored (%)	91
Ramachandran outliers (%)	3.3
Clash score	11.03
Average B-factor	79.70

Table 4.5 AMP bound Ct-Rio2 Scalepack Logfile at pH 7. The Ct-Rio2 crystals were formed in the presence of ADP/$BeCl_2$/KF in the hanging drop. The crystals were produced in 0.1M HEPES pH 7.0. 30% v/v jeffamine ED-2001pH 7.0 at the 2:1 protein to well solution ratio.

	Overall	InnerShell	OuterShell
Low resolution limit	59.49	59.49	3.44
High resolution limit	3.26	10.31	3.26
Rmerge (within I+/I-)	0.110	0.020	0.907
Rmerge (all I+ and I-)	0.127	0.025	1.027
meas (within I+/I-)	0.141	0.026	1.160
Rmeas (all I+ & I-)	0.143	0.028	1.154
Rpim (within I+/I-)	0.087	0.016	0.717
Rpim (all I+ & I-)	0.065	0.013	0.520
Rmerge in top intensity bin	0.021	-	-
Total number of observations	28637	902	4148
Total number unique	5955	188	862
Mean((I)/sd(I))	13.0	56.4	1.7
Mn(I) half-set correlation CC(1/2)	0.996	0.999	0.564
Completeness	99.9	98.9	100.0
Multiplicity	4.8	4.8	4.8
Anomalous completeness	93.8	95.4	92.9
Anomalous multiplicity	2.3	2.4	2.3
DelAnom correlation between half-sets	0.058	0.200	0.023
Mid-Slope of Anom Normal Probability	1.063	-	-
Average unit cell	119	119	72.35
Space group	H3		
Average mosaicity:	0.16		

Table 4.6 AMP-bound Ct-Rio2 phenix.xtriage logfile at pH 7. The Ct-Rio2 crystals were formed in the presence of ADP/$BeCl_2$/KF in the hanging drop. The crystals were produced in 0.1M HEPES pH 7.0. 30% v/v jeffamine ED-2001 pH 7.0 at the 2:1 protein to well solution ratio.

Res. Range	I/sigI>1	I/sigI>2	I/sigI>3	I/sigI>5	I/sigI>10	I/sigI>15
59.49 - 7.69	98.3%	97.6%	97.0%	69.5%	0.0%	0.0%
7.69 - 6.10	98.4%	96.7%	91.3%	63.7%	0.0%	0.0%
6.10 - 5.33	97.7%	93.7%	84.8%	50.1%	0.0%	0.0%
5.33 - 4.84	97.8%	91.8%	82.0%	43.2%	0.0%	0.0%
4.84 - 4.50	98.3%	90.9%	77.7%	41.6%	0.0%	0.0%
4.50 - 4.23	97.8%	86.9%	73.8%	38.0%	0.0%	0.0%
4.23 - 4.02	96.7%	84.5%	67.3%	26.8%	0.0%	0.0%
3.84 - 3.70	96.1%	73.7%	53.4%	14.3%	0.0%	0.0%
3.70 - 3.57	93.8%	68.2%	44.1%	13.2%	0.0%	0.0%
3.57 - 3.46	90.6%	53.2%	30.7%	6.7%	0.0%	0.0%
3.46 - 3.36	87.9%	43.4%	22.1%	3.0%	0.0%	0.0%
3.36 - 3.27	81.8%	38.9%	18.6%	1.5%	0.0%	0.0%
3.27 - 3.19	80.7%	33.0%	15.9%	2.1%	0.0%	0.0%

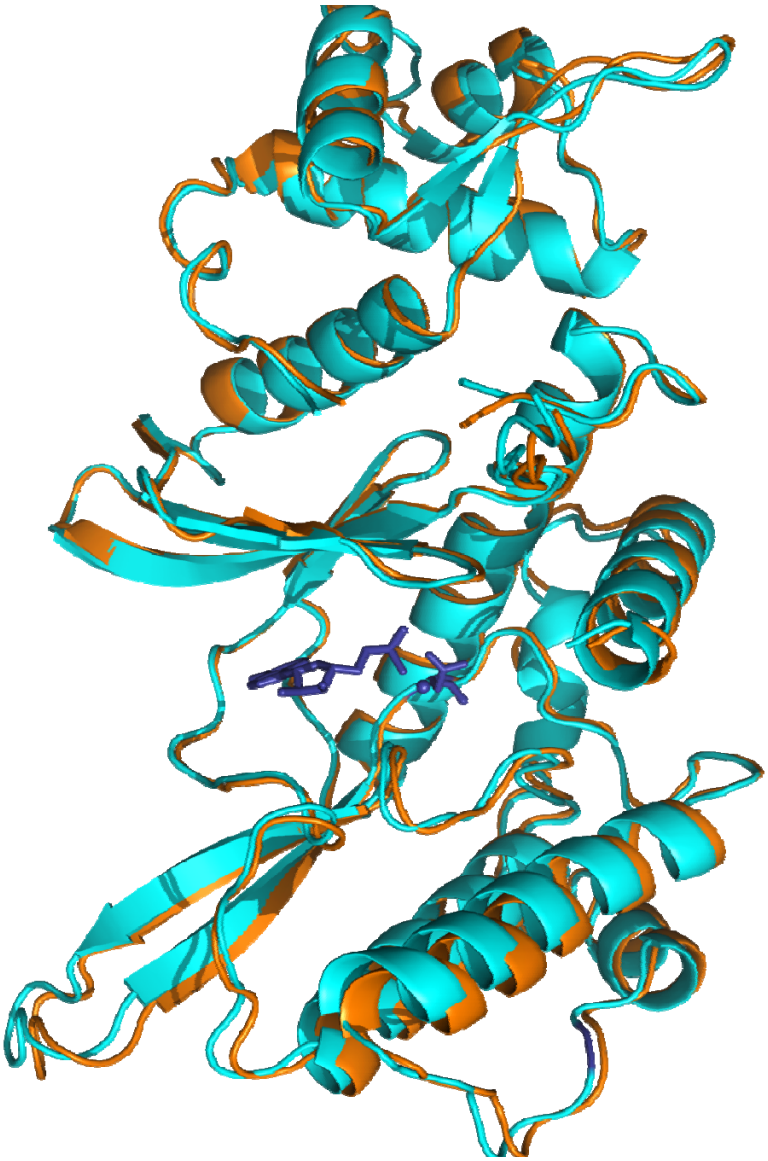

Figure 4.6 Alignment between bound AMP to CtRio2 at both pH 5.2 and pH 7.0: The structures of Ct-Rio2 are aligned based on the AMP location. The orientation of Asp 257 is the same for both of the crystals.

Figure 4.7 The Ct-Rio2 crystals formed in the presence of ADP/Na_2VO_4 in the hanging drop at pH 5.4: The crystals were produced with 0.05M ammonium acetate, 0.01 magnesium chloride, 0.05M TRIS hydrochloride pH 7.5, 10% v/v 2-methyl-2,4-pentandiol.

Table 4.7 Data collection and current refinement statistics for AMP-bound Ct-Rio2 at the pH 5.4: The Ct-Rio2 crystals formed in the presence of ADP/Na_2VO_4 in the hanging drop at pH 5.4: The crystals were produced with 0.05M ammonium acetate, 0.01 magnesium chloride, 0.05M TRIS hydrochloride pH 7.5, 10% v/v 2-methyl-2,4-pentandiol.

Resolution range (Å)	59.2 - 2.07 (2.144 - 2.07)
Space group	R 3: H
Unit cell	118.402 118.402 72.43 90 90 120
Total reflections	56005 (1547)
Unique reflections	23038 (809)
Multiplicity	3.6 (1.9)
Completeness (%)	99.93 (99.65)
Mean I/sigma (I)	5.80 (0.65)
Wilson B-factor	50.83
R-merge	0.7257 (1.091)
R-meas	0.8069
CC1/2	0.99 (0.441)
CC*	0.997 (0.782)
R-work	0.2017 (0.2774)
R-free	0.2511 (0.3206)
Number of non-hydrogen atoms	2567
Macromolecules	2543
Ligands	24
Protein residues	338
RMS (bonds)	0.014
RMS (angles)	1.59
Ramachandran favored (%)	93
Ramachandran outliers (%)	2.1
Clashscore	4.82
Average B-factor	77.00

Table 4.8 AMP-bound Ct-Rio2 phenix.xtriage logfile at the pH 5.4: The Ct-Rio2 crystals formed in the presence of ADP/Na_2VO_4 in the hanging drop at pH 5.4: The crystals were produced with 0.05M ammonium acetate, 0.01 magnesium chloride, 0.05M TRIS hydrochloride pH 7.5, 10% v/v 2-methyl-2,4-pentandiol.

Res. Range	I/sigI>1	I/sigI>2	I/sigI>3	I/sigI>5	I/sigI>10	I/sigI>15
59.24 - 4.28	99.9%	99.8%	99.5%	98.0%	79.3%	51.3%
4.28 - 3.39	99.9%	99.6%	99.0%	97.2%	88.4%	67.9%
3.39 - 2.97	99.5%	97.7%	95.8%	91.3%	72.8%	42.3%
2.97 - 2.69	99.1%	93.7%	89.0%	78.5%	47.2%	19.8%
2.69 - 2.50	97.1%	85.5%	76.3%	59.4%	24.7%	8.0%
2.50 - 2.35	94.4%	76.8%	63.1%	41.8%	10.1%	2.0%
2.35 - 2.24	90.8%	65.8%	48.0%	24.3%	3.8%	0.6%
2.24 - 2.14	87.6%	52.1%	31.4%	11.5%	1.0%	0.1%
2.14 - 2.06	80.1%	34.2%	15.1%	3.1%	0.2%	0.0%
2.06 - 1.99	74.0%	22.8%	7.5%	1.1%	0.1%	0.0%
1.99 - 1.92	66.3%	13.9%	2.8%	0.6%	0.0%	0.0%
1.92 - 1.87	59.8%	9.1%	1.0%	0.0%	0.0%	0.0%
1.87 - 1.82	56.5%	5.8%	0.5%	0.0%	0.0%	0.0%
1.82 - 1.77	49.3%	4.2%	0.3%	0.0%	0.0%	0.0%

The completeness of data for which I/sig(I)>3.00, exceeds 85% for resolution ranges lower than 2.69A.

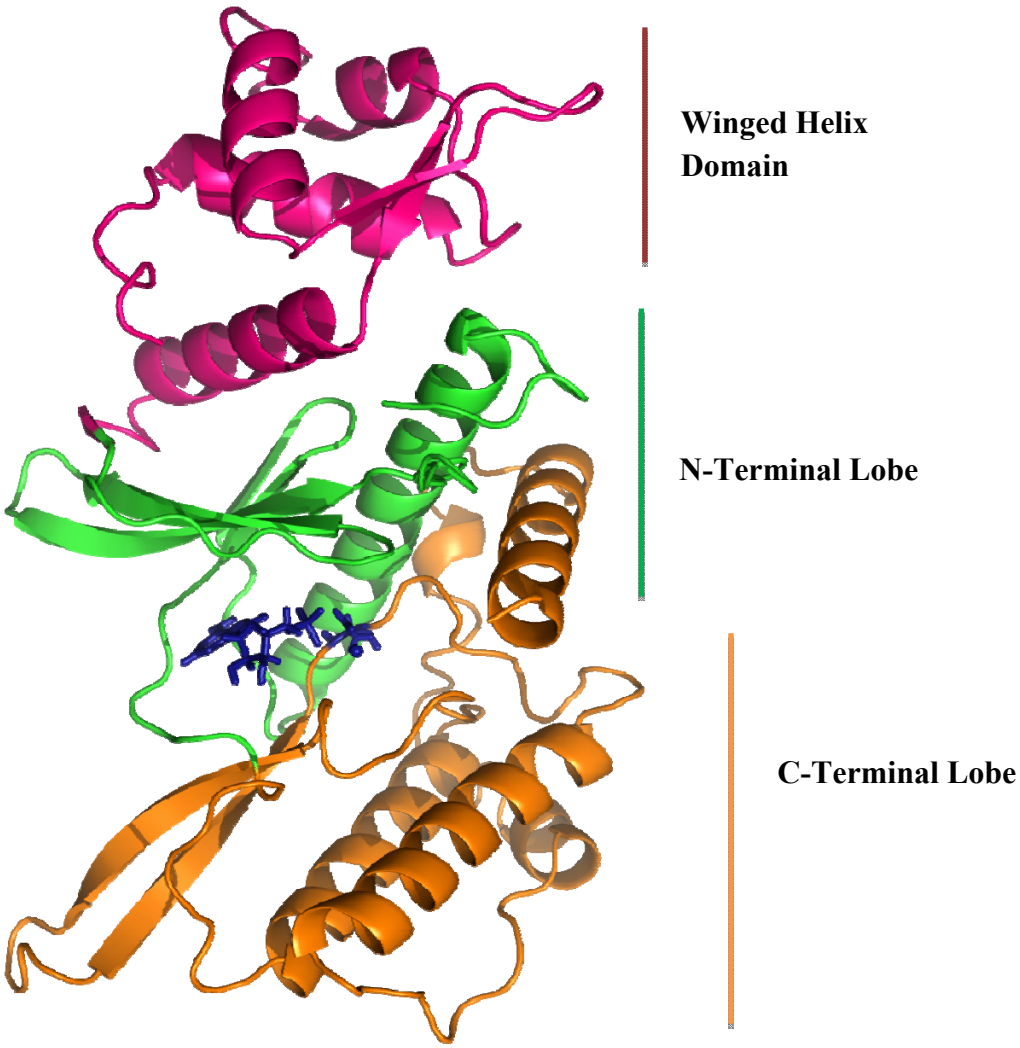

Figure 4.8 Crystal Structure of AMP bound Ct-Rio2: The crystal structure of Ct-Rio2 shows a bilobal kinase domain (green and orange) connected to an N-terminal winged helix domain (pink). The "hinge" region connects the N- and C- terminal lobes of the kinase domain.

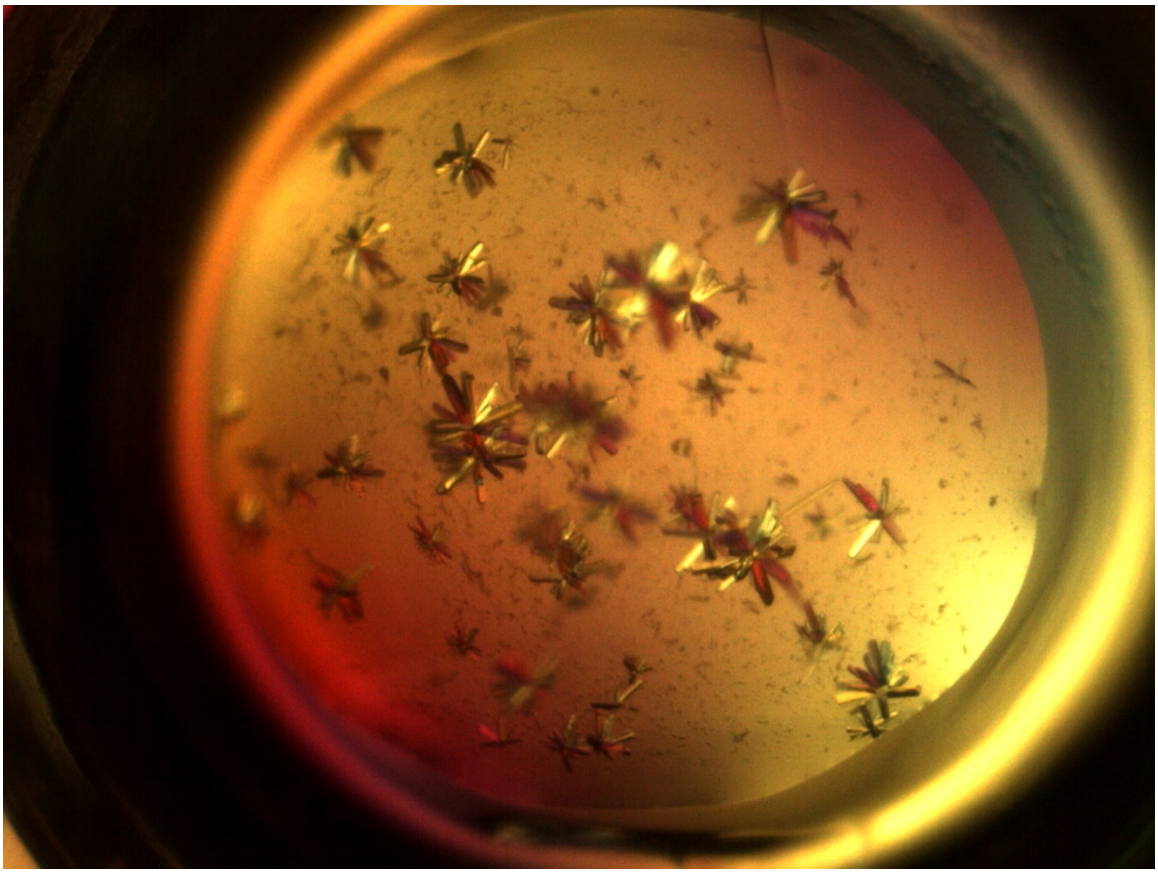

Figure 4.9 The Ct-Rio2 crystals formed in the presence of ADP/Na_2VO_4 in the hanging drop at pH 7.0: The crystals were produced in 0.05M ammonium acetate, 0.01 magnesium chloride, 0.05M TRIS hydrochloride pH 7.5, 10% v/v 2-methyl-2,4-pentandiol.

Table 4.9 AMP-bound Ct-Rio2 Scalepack Logfile at pH 7. The Ct-Rio2 crystals formed in the presence of ADP/Na_2VO_4 in the hanging drop at pH 5.4: The crystals were produced in 0.05M ammonium acetate, 0.01 magnesium chloride, 0.05M TRIS hydrochloride pH 7.5, 10% v/v 2-methyl-2,4-pentandiol.

Resolution range (Å)	59.48 - 3.252 (3.369 - 3.252)
Space group	R 3: H
Unit cell	119.252 119.252 72.75 90 90 120
Total reflections	22654 (1412)
Unique reflections	6037 (522)
Multiplicity	3.8 (2.7)
Completeness (%)	99.47 (99.84)
Mean I/sigma (I)	1.12 (0.13)
Wilson B-factor	89.46
R-merge	0.8686 (3.86)
R-meas	0.9781
CC1/2	0.967 (0.477)
CC*	0.992 (0.804)
R-work	0.1737 (0.3155)
R-free	0.2654 (0.2951)
Number of non-hydrogen atoms	2731
Macromolecules	2695
Ligands	36
Protein residues	343
RMS (bonds)	0.013
RMS (angles)	1.70
Ramachandran favored (%)	86
Ramachandran outliers (%)	3.6
Clashscore	18.86
Average B-factor	94.40

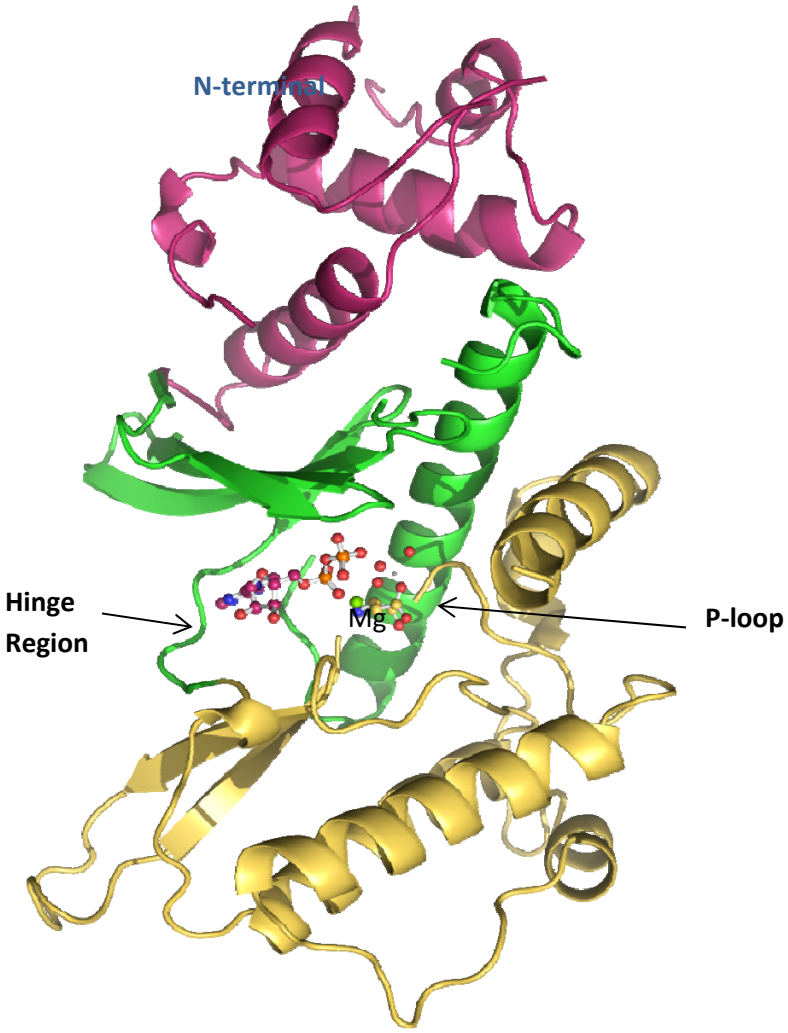

Figure 4.10 Crystal structure of Ct-Rio2 bound to ADP and transition state analogue sodium orthovanadate: The crystal structure of Ct-Rio2 shows a bilobal kinase domain (green and yellow) connected to an N-terminal winged helix domain (pink). The "hinge" region connects the N- and C- terminal lobes of the kinase domain. The ADP binding loop "P-loop", metal binding loop used to bind Mg, and catalytic loop which contains the catalytic D232 are indicated by arrows. The cartoon representation was made by PYMOL.

Chapter 5: Structural Analysis of AMP- and ADP-Bound Ct-Rio2

5.1 Introduction

In the last chapter, I discussed how Ct-Rio2 was purified and how the X-ray diffraction data were acquired. In this chapter, I will analyze the structural features of Ct-Rio2. The first structure of eukaryotic Rio2 kinases has been found in our lab in Ct-Rio2. This protein is significantly extended compared to the archaeal version of Rio2 for which we currently have structural information.

5.2 Obtaining the Crystal Structure of Ct-Rio2

Both apo-ctRio2 crystals and Ct-Rio2 with ATP crystals were solved by molecular replacement using the known structure of afRio2 with ADP and Mn^{2+} bound. Vatsala Sagar provided in her study what the apo and ATP-bound Ct-Rio2 looks like. Using the purification techniques that were optimized in my work, Ct-Rio2 was crystalized in in the presence of ADP and AMP. Our initial goal was to obtain the transition state analogues of Ct-Rio2 in presence of $BeCl_2$/KF and Na_2VO_4. These structures could be interesting since they represent the active site scheme of the transfer of Phosphate group from ATP to Asp residue on Ct-Rio2.

A previous model of Ct-Rio2 was used for molecular replacement. The program used for molecular replacement was the Phaser crystallography software suite in PHENIX. Additionally, Refinement was done with phenix.refine in Phaser. Rfree was monitored by using 10% of the reflections as a test set. The crystallographic refinement statistics are shown in last chapter along with few figures that depicted the structures of Ct-Rio2 using PYMOL.

5.3 Crystal Structure of Ct-Rio2 with ADP, $BeCl_2$/KF

Crystals of the Ct-Rio2 structure and the ADP, $BeCl_2$ and KF complex were isomorphous and belonged to the space group H3, with one molecule per asymmetric unit. Diffraction of the complex was measured on a synchrotron source (NE-CAT, Argonne National Laboratories Advanced Photon Source). Although Ct-Rio2 was crystallized in the presence of 20 mM ADP and 2.5 mM MgCl2 and $BeCl_2$ and KF, no ADP was observed in the active site of the enzyme. Details of the crystallization methods were described in Chapter 4. We thought this could have happened because the ADP was degraded or that it was not bound to the protein in the crystals. We also didn't observe the Beryllium complexes in the crystal.

However, as discussed in the last chapter, the AMP was found in the crystal in both conditions when the $BeCl_2$, KF, and ADP were added at both pH of 5.4 and pH of 7. This observation especially at pH 7 can mean that Ct-Rio2 can be capable of performing ADPase activity as well as being an ATPase. This hypothesis could be less probable that meaning that the ADP molecule is not stable in presence of $BeCl_2$ and KF. The latter is more convincing since we have observed previously in the ATP bound protein and in the last chapter when vanadate was used that ADP can bind to the protein and is not degraded at pH 7.

5.4 The Structural Analysis of the AMP-Bound Ct-Rio2

The structural features of the AMP-bound Ct-Rio2 were very similar to what has been observed previously in the ATP-bound Ct-Rio2 structure (**Figure 5.1**). The N-terminus of Ct-Rio2 bound to the AMP contains a winged helix domain. It is comprised of four α helices followed by a β-strand wing β-strand motif, and another α helix. The Rio domain in Ct-Rio2 consists of an N-terminal lobe and a C-terminal lobe with a flexible hinge region between them. AMP binds in the large cleft between the two lobes (**Figure 5.2**).

The N-terminal lobe consists of a five-stranded antiparallel β sheet and a long α helix. The C-terminal lobe of the RIO domain consists of two antiparallel β strands and two α helices. The third α helix seen in af-Rio2 is missing in the Ct-Rio2 structure. However, instead there is another C-terminal α helix that is not seen in the af-Rio2 structure partially obstructing the opening to the active site. This α helix, labeled αJ, is situated in the cleft between the C and N terminal lobes of the RIO kinase domain.

The active site of Ct-Rio2 contains three loops—the P-loop, the "DFG" loop, and the catalytic loop. The P-loop is involved in binding AMP and orienting the phosphate in a position to be attacked by the nucleophilic serine in a phosphoryl transfer reaction. The "DFG" loop binds and positions the metal ions that coordinate the phosphate oxygens of AMP and help orient the nucleotide. The catalytic loop contains the conserved N and D residues necessary for catalysis and metal binding. In typical protein kinases the sequence of the P-loop is GXGXXG. In Ct-Rio2 the sequence is GVGKES. The sequence of the DFG loop in Ct-Rio2 is IDFPQ and the catalytic loop sequence is DLSQYN (**Figure 5.3**).

Polar contacts could be observed between the α-phosphate of AMP and Mg2+ with Met166, Ser108, Asp257 and Asn234. This structure was perfectly fit in the 2Fo-Fc map of Ct-Rio2 (**Figure 5.4**).

5.5 Crystal Structure of Ct-Rio2 in the Presence of ADP and Na_2VO_4

For the Ct-Rio2/ADP/Vanadate complex, we obtained the diffraction data sets at 3.25 Å. Data was indexed, integrated and scaled using HKL2000 package program. The crystals belonged to R3 space group and the unit cell dimensions were a= 119.252 Å, b= 119.252 Å, c= 72.75 Å and α=90.000, β=90.000, γ=120.000. Crystal structure of Ct-Rio2 at pH 7 shows the vanadate and ADP in the active site bound to the Asp257 and Mg^{2+}. The Asp257 seems to have

polar interaction to both VO_4 and Mg^{2+}. The overlay of the active site of Ct-Rio2 bound to ADP and Orthovanadate vs. Ct-Rio2 bound to AMP displays a dramatic movement of Asp257 from the AMP bound complex to the vanadate bound complex (**Figure 5.5-5.6**).

5.6 Summary

The crystal structure of Ct-Rio2 was determined and the structures of AMP and ADP/VO_4 bound Ct-Rio2 were compared. In both of the structures the Ct-Rio2 protein showed a bilobal structure that was expected. In the AMP bound protein the Asp257 bonded the α-phosphate of the AMP in a different location and orientation compared to that of the ADP bound protein.

The movement of asp257 in the transition state analog structure reveals that in the presence of the ATP the protein bilobal structure opens up to accommodate the ligand. The orientation of Asp257 is then changed to bind the gamma-phosphate of the ATP. In this case the vanadate structure mimics the γ-phosphate of the ATP. The structure of Ct-Rio2 in a complex with AMP aligns well with the previous apo protein structure. Moreover, Ct-Rio2 as expected showed an extended C-terminus containing αJ. The αJ interacts with the kinase domain through both hydrophobic and hydrophilic interactions with αC, αF, the flexible loop and the metal binding loop.

The Asp257 displays to both VO_4 and Mg^{2+} through polar interactions. In the bilobal kinase domain of Ct-Rio2 at pH 5.4, AMP is bound with no trace of VO_4 in its active site. The winged helix domain of the two structures is not completely superimposed. The overlay of the AMP region in the crystal structures of Ct-Rio2 bound to AMP and ADP at two different pH showed that both C and N terminal lobes were apart. The overlay of the active site of Ct-Rio2

bound to ADP and Orthovanadate vs. Ct-Rio2 bound to AMP displays a dramatic movement of Asp257 from the AMP bound complex to the vanadate bound complex.

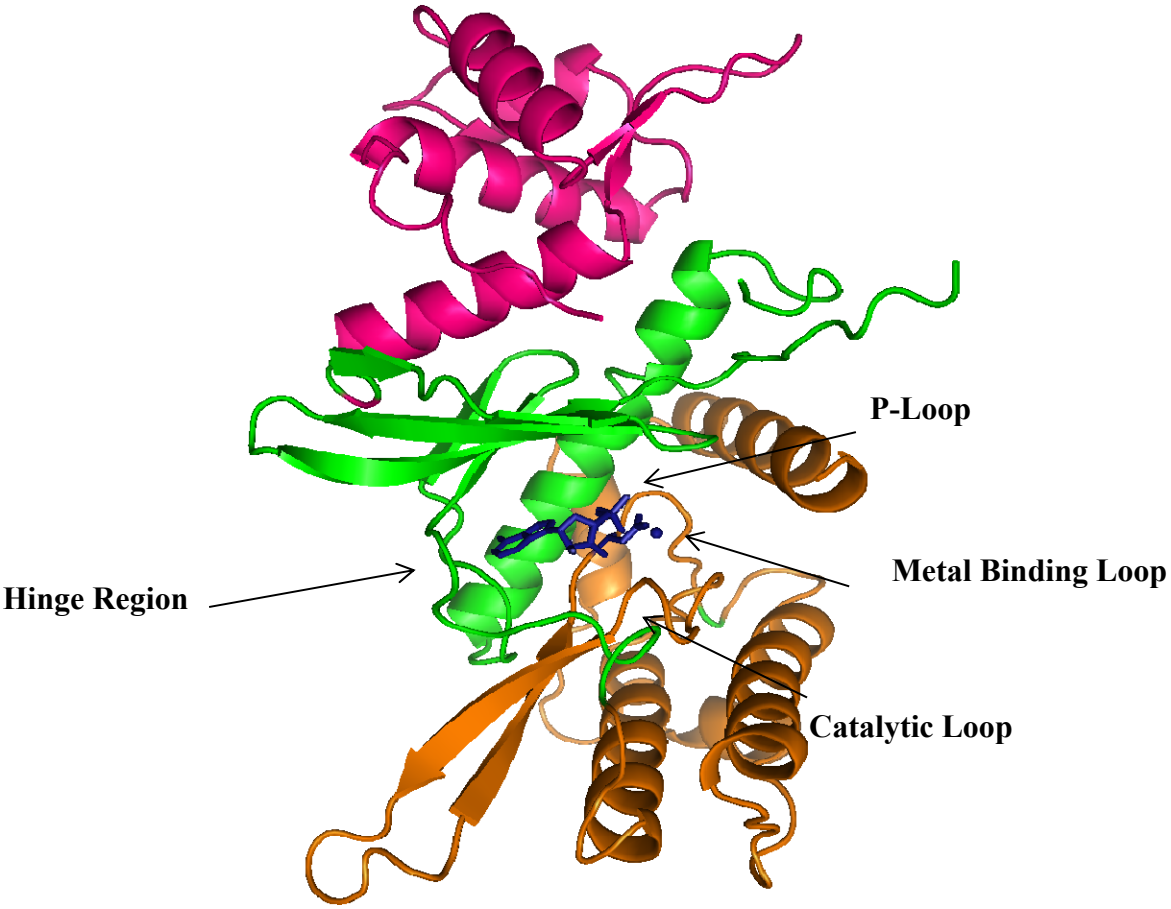

Figure 5.1 Crystal Structure of AMP bound Ct-Rio2 shows bilobal structure: The crystal structure of Ct-Rio2 shows a bilobal kinase domain (green and orange) connected to an N-terminal winged helix domain (pink). The "hinge" region connects the N- and C-terminal lobes of the kinase domain. The ATP binding loop "P-loop", metal binding loop used to bind Mg, and catalytic loop which contains the catalytic D257 are indicated by arrows. The cartoon representation was made by PYMOL.

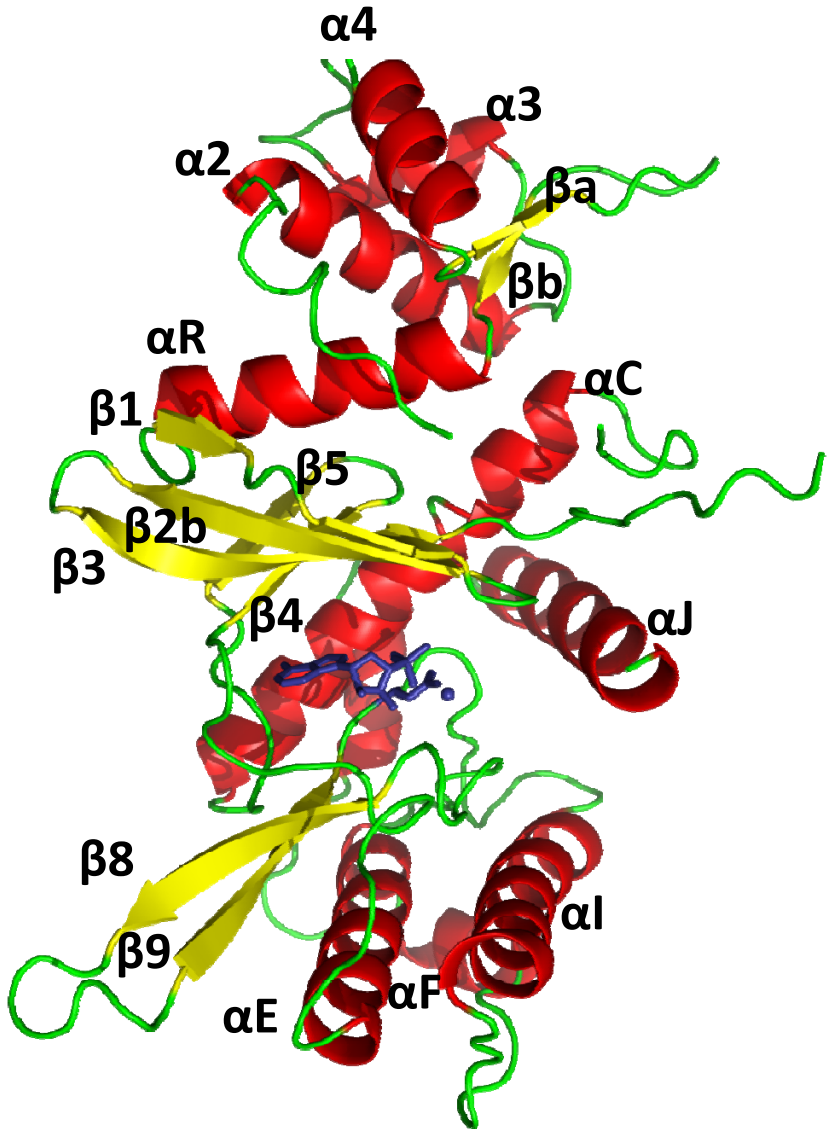

Figure 5.2 Crystal Structure of AMP-bound Ct-Rio2: The N-terminus of Ct-Rio2 contains a winged helix domain. It is comprised of four α helices followed by a β-strand wing β-strand motif, and another α-helix. AMP binds in the large cleft between the two lobes. The N-terminal lobe consists of a five-stranded antiparallel β sheet and a long α helix. The C-terminal lobe of the Rio domain consists of two antiparallel β strands and two α helices.

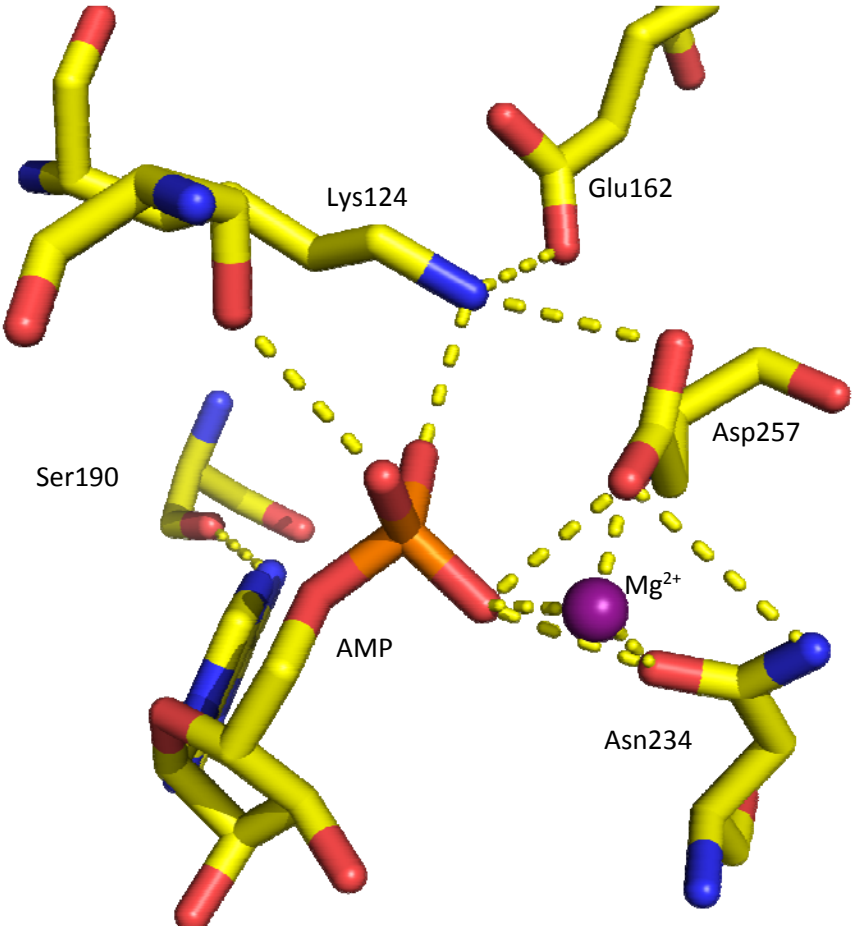

Figure 5.3 Polar Contacts with the AMP: Crystal structure of AMP bound Ct-Rio2 shows that the α phosphate of AMP makes polar contacts with Met 166, Asp 257, Asn 234 and Mg ion.

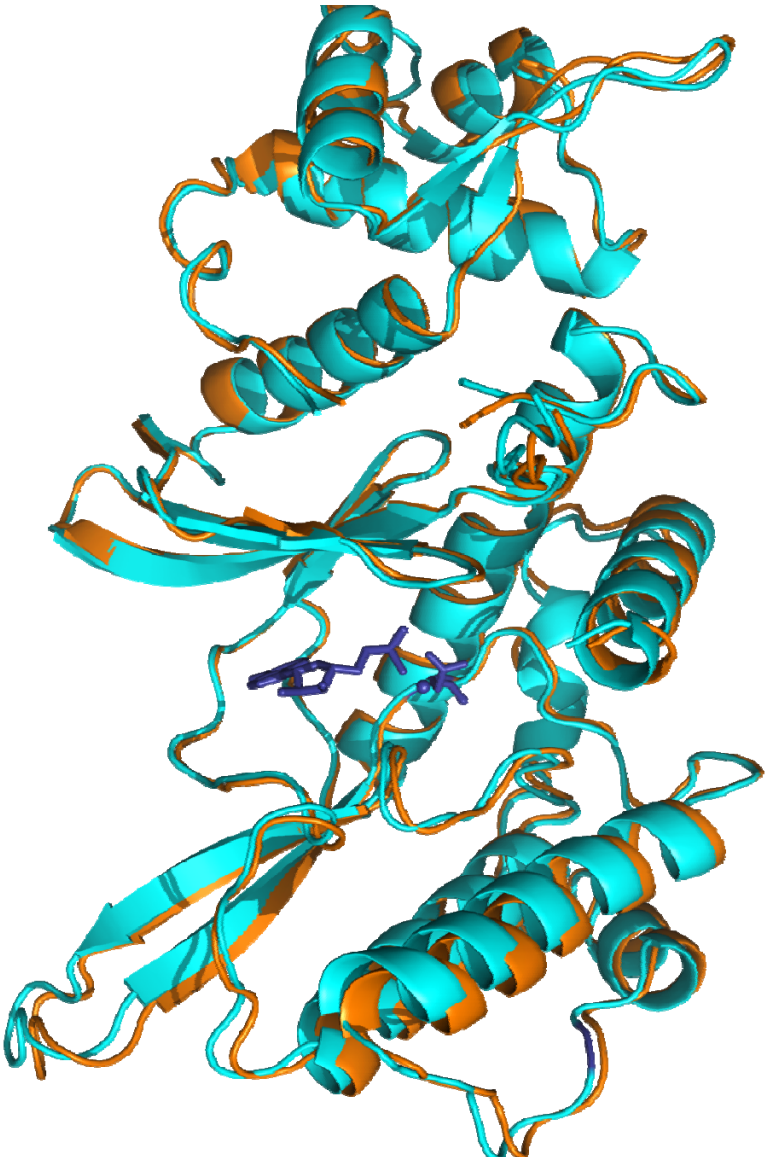

Figure 5.4 Alignment between bound AMP to CtRio2 at both pH=5.2 and pH=7.0: The structures of Ct-Rio2 are aligned based on the AMP location. The orientation of Asp 257 is the same for both of the crystals.

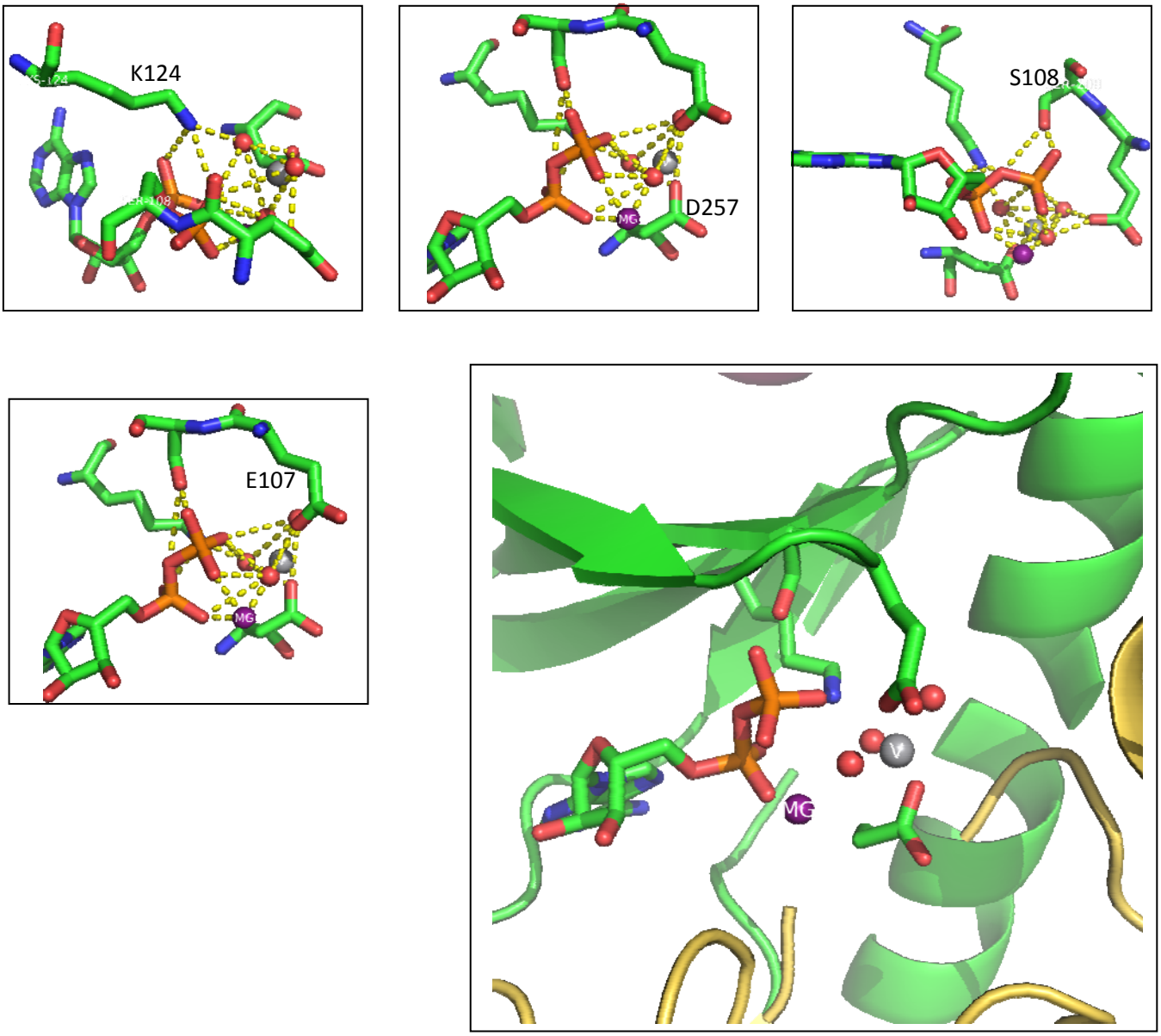

Figure 5.5 The active site of Ct-Rio2 bound to ADP and Orthovanadate: The crystal structure of Ct-Rio2 with vanadate and ADP shows that polar Contacts can be observed between α and β phosphate of ADP with Lys. The β phosphate of ADP is simultaneously bound to vnadate oxygens, $Mg^{2+}$and Asp257.

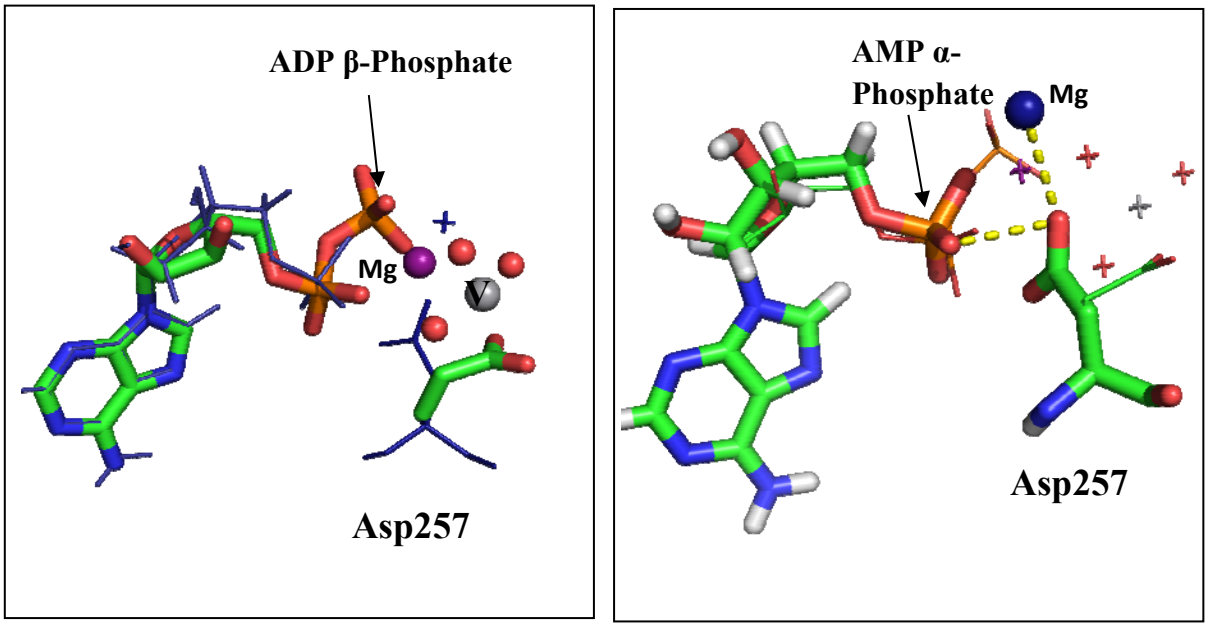

Figure 5.6 The overlay of the active site of Ct-Rio2 bound to ADP and Orthovanadate vs. Ct-Rio2 bound to AMP: The Asp 257 displays a dramatic movement from the AMP bound complex to the vanadate bount complex.

Chapter 6: Evolution of the Eukaryotic Protein Kinases

6.1 Evolutionary Origins of the Eukaryotic Protein Kinases

For many years after the discovery of protein phosphorylation, the phosphorylation of proteins on serine, threonine, and tyrosine residues was a phenomenon that was thought to be restricted to the eukaryotes, in which these modifications perform important regulatory functions[78]. The analogous regulatory role in bacteria was attributed to the sensor kinases and the phosphoenolpyruvate-dependent phosphotransferase systems, which typically catalyze the phosphorylation of proteins on histidine and aspartate residues, respectively. Later on the observation of proteins containing phosphoserine, phosphothreonine, or phosphotyrosine residues in bacteria and archaea illustrated that several mechanisms of Ser/Thr or Tyr protein phosphorylation have been recognized in prokaryote[79].

6.2 The Rio Family

In order to examine the evolutionary origins of the eukaryotic protein kinase (ePK) superfamily, extensive analysis has been conducted over the past years on the proteins encoded by the completely sequenced bacterial and archaeal genome[80]. Five distinct families of putative protein kinases are detected with representatives in bacteria and archaea that share a common ancestry with the eukaryotic protein kinases. Except Rio proteins the other four of these protein families have not been identified previously as protein kinases. From the phylogenetic distribution of these families, the existence of an ancestral protein kinase(s) prior to the divergence of eukaryotes, bacteria, and archaea is predicted[81]. Most of the eukaryotes and archaea contain two paralogs of the Rio family. These proteins form two (or possibly three) distinct subfamilies of paralogs. These protein kinases along with other phosphatases catalyze

the reversible protein phosphorylation which is mainly responsible for cellular regulation and signal transduction in all three domains of life. Since any abnormal levels of phosphorylation is known to be responsible for variety of sever diseases, there's a considerable therapeutic promise in obtaining a detailed understanding of phosphorylation processes, both within a specific cell and in their evolutionary context[80].

6.3 The Common Classifications of Protein Kinases

Based on the KinBase resource, protein kinases are classified into two broad groups of conventional eukaryotic protein kinases (ePKs) and atypical protein kinases (aPKs)[80]. By examining the sequences and the catalytic domains and considering the known modes of regulation, ePKs are subdivided into eight families. The eight ePK families are: the AGC family, the CAMKs, the CK1 family, the CMGC family, the RGC family, the STE family, the TKfamily and the TLK family[50, 82].

The aPKs are a smaller set of protein kinases that have shown the protein kinase activity but do not share sequence similarities with ePKs. The aPKs are comprised of the following subfamilies: alpha, PIKK, PHDK, RIO, BRD, ABC1 and TIF1. In order to examine the evolutionary origins of the eukaryotic protein kinase superfamily, Koonin et al have conducted an extensive analysis of the proteins encoded by the completely sequences bacterial and archaeal genomes. From the phylogenetic distribution of the resulted protein families, the existence of an ancestral protein kinase is proposed that existed prior to divergence of eukaryotes, bacteria and archaea.

Rio kinases were originally identified in Archaea and Eukarya and a common ancestor with an ancestral Rio gene has been predicted. Phylogenetic analysis has revealed that Rio kinases are distributed in all three domains of life. The Rio kinases are classified into four groups

including: Rio1, Rio2, Rio3 and RioB. Rio2 exist in archaea, eukaryotes and bacteria with high sequence and activity similarities. Canonical protein kinases are only observed in archaea. Canonical protein kinases are very similar to Rio2 proteins but are more active in ATPase and autophosphorylation activity[79].

Computational studies have revealed that the arginine residue at position 228 highly affects the activity of canonical protein kinases. This residue in Rio family is replaced by either alanine or glycine.

In this study, we are interested to study mutation at 228 of Rio2, and find out how this mutation would affect the structure and activity of the Rio2 family. We are interested to investigate the evolutionary relationship between Rio kinases and the canonical eukaryotic protein kinases. Our hypothesis is that canonical eukaryotic protein kinases have evolved from Rio kinases. We have made several mutations in a key differentiating position of the kinase domain sequence of Ct-Rio2, G228, and have collected biochemical data to test the ability of the mutated Rio kinase to autophosphorylate and carry out ATP hydrolysis.

6.4 Purification of Rio2 Mutants from *Chaetomium Thermophilum*

Following the optimized procedure outlined in chapter 2 Ct-Rio2 mutants were designed and were further expressed, purified and crystallized. The p7-Ct-Rio2 vectors were transformed into *E. coli* Rosetta^{TM}(DE3) pLysS (Novagen), and were used to perform test expression of the cells before purifying large amount of protein. The result of test expression gel revealed that cells with the mutants could produce a band of Ct-Rio2 at 44.6kDa (**Figure 6.1-6.2**).

6.5 Crystallization of Ct-Rio2 Mutants

We have also co-crystallized Ct-Rio2 G228 point mutants with ATP in order to determine the structural differences between the mutant and wild-type. Hence, following the

same procedure that was developed for wild type Ct-Rio2, five mutants of Ct-Rio2 were purified and co-crystalized with ATP/Mg^{2+}. These mutants included G228A, G228D, G228N, G228R, and G228V. While it was difficult to obtain any crystals for G228R and G228V, several crystals were formed for the G228A, G228D, and G228N mutants and x-ray diffraction data sets were collected (**Figure 6.3**).

The best crystals for G228A belonged to the Wizard screen condition number 70 (10% (v/v) 2-propanol, 0.1M imidazole pH 8.0). For G228D the best hit belonged to PEG condition 40 (0.2 M Sodium phosphate dibasic dehydrate, 20% w/v Polyethylene glycol 3,350) and for G228N belonged to Wizard screen number 66 (20% (w/v) PEG-3000. 0.1M Tris pH 7.0, 0.2M $Ca(OAc)_2$). The crystal structure of Ct-Rio2 with Asn mutation at residues 228 shows a bilobal kinase domain (blue and yellow) connected to N-terminal winged helix domain (green). The "hinge" region connects the N- and C- terminal lobes of the kinase domain.

6.6 Crystal Structure of AMP Bound G228N

Data was indexed, integrated and scaled using HKL2000 package program. The crystals belonged to R3 space group and the unit cell dimensions were a=118.623 Å, b=118.623 Å, c=72.556 Å and α=90.000, β=90.000, γ=120.000 (**Table 6.1**). The structure was solved using PHENIX. Subsequent rounds of model building were carried out in COOT accompanied by several rounds of refinement using phenix-refine. Waters were added to the structure using PHENIX. The multiplicity was 5.2 and completeness was 99.71. The final model had a Rwork/Rfree of 20.03/24.69 (**Table 6.1**).

6.7 Crystal Structure of AMP Bound G228A

Data was indexed, integrated and scaled using HKL2000 package program. The crystals belonged to R3 space group and the unit cell dimensions were a=119.658 Å, b=119.658 Å,

c=72.342 Å and α=90.000, β=90.000, γ=120.000 (**Table 6.2**). The structure was solved using PHENIX. Subsequent rounds of model building were carried out in COOT accompanied by several rounds of refinement using phenix-refine. Waters were added to the structure using PHENIX. The multiplicity was 2.5 and completeness was 98.10. The final model had a Rwork/Rfree of 18.65/25.78 (**Table 6.2**).

6.8 Crystal Structure of AMP Bound G228D

Data was indexed, integrated and scaled using HKL2000 package program. The crystals belonged to R3 space group but the twining was high. Tried deleting few frames in iMOSFLM but couldn't solve the twining issue.

6.9 Results and Discussion

The crystal structure of Ct-Rio2 G228N, and G228A showed a bilobal kinase domain connected to a N-terminal winged helix domain (**Figure 6.4-6.5**).

The overlay of the C-Terminal domain in the crystal structures of G228N and G228A of Ct-Rio2 showed that the winged region was not superimposed perfectly and the presence of the Asn at the 228 opened up the phosphate binding loop (**Figure 6.6**). Upon the overlay of the bound AMP groups none of the C and N terminal lobes neither the winged helix domain of the structures was completely superimposed (**Figure 6.7**). The overlay of the c-terminal region in the crystal structures of G228N, G228A and G228 of Ct-Rio2 bound to AMP showed that while the A and wild type Glycine Ct-Rio2 were superimposable the Asn mutated Ct-Rio2 was structurally different.

6.10 ATPase Activity and Autophosphorylation Comparison

Based on our steady state hydrolysis assay results, we have found that the proteins with different point mutations at 228, show very different activities in their ATPase and Auto-

phosphorylation experiments (**Figure 6.8**). The G228A and Ct-Rio2 with G228 are similarly active in the hydrolysis of ATP but mutants with larger residues such as G228V are less active. The mutants with charged residues such as those we find mostly in the EPKs[5] including G228D and G228N and G228R are almost inactive in ATP hydrolysis. In line with our predictions, the auto-phosphorylation of G228D and G228N are less pronounced than the other mutations. Our crystal structure finds very similar structure for the G228A and G228V mutants to that of Ct-Rio2 (**Figure 6.9**).

6.11 Summary

The procedure that was used for the purification of the mutants was at the pH of 5.4. We believe that the ATP that was originally mixed with the protein for crystallization has degraded to AMP. But the comparison of the crystal structure of all of the mutants can be studied since all of the mutants were bound to AMP. When I overlay the structures based on the AMP in their active site or based on the C terminus, the mutant with N at the residue 228 had an active site that was wider than that of Glycine and Alanine crystals. This opening of the catalytic loop seems to decrease the ATPase activity and at the same time increase the kinase activity and autophosphorylation. The same phenomena are observed in the epks. In the canonical epks the active site is wide due to the arginine residue at 228. We have observed in our ATPase and autophosphorylation experiments that residues that have R or a residue as large as arginine are more capable of autophosphorylation. However, the alanine and glycine residues are better as ATPases than Kinases. Since the Rio kinases have existed 1.5 billion years before the emergence of the cananoical eukaryotic protein kinases, they are very likely to have been the ancestors of the epks. This hypothesis is strongly supported based on the sequence alignment studies. This work suggest that by mutation of the residue 228 epks have switched from a good ATPase kinase

to a better autophosphorylation kinase. The core of the epks is very similar to Rio kinases and many residues are conserved. The few changed residues and the secondary structures that are later added to the canonical protein kinases could have a great impact on their function and activity as kinases.

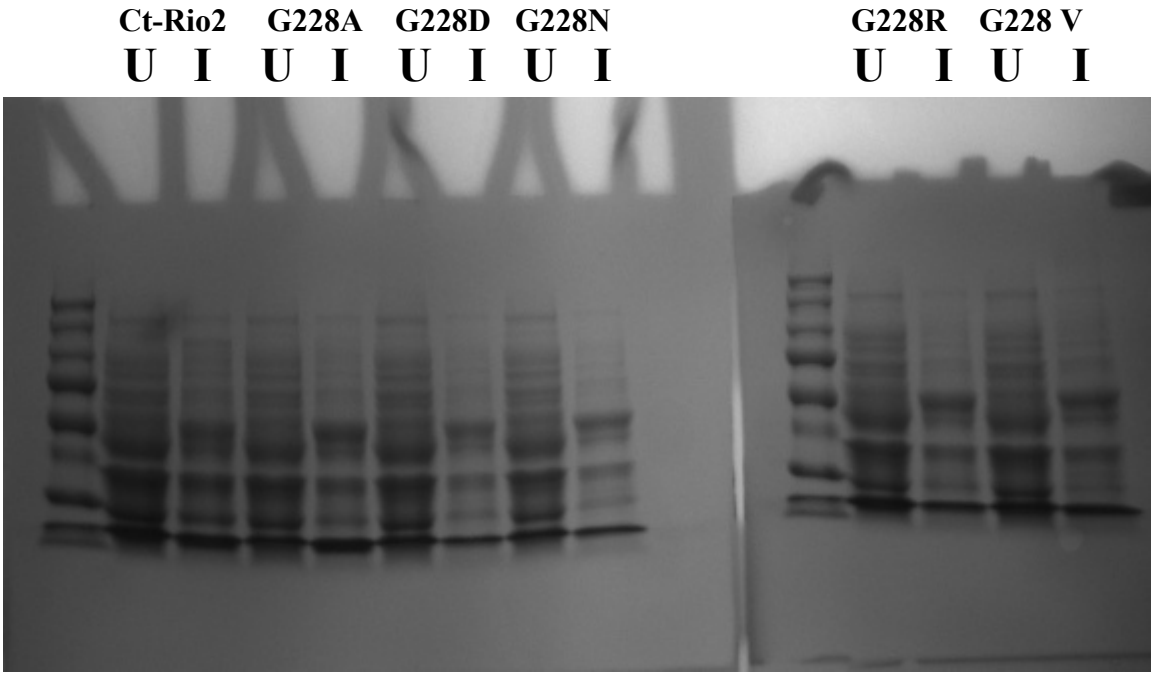

Figure 6.1 Testing of the expression of Ct-Rio2 protain and its G228 mutants in E.Coli: The induced (I) samples show a single band at 46 kD which corresponds to Ct-Rio2 and its mutant. The uninduced (U) samples didn't display the same band.

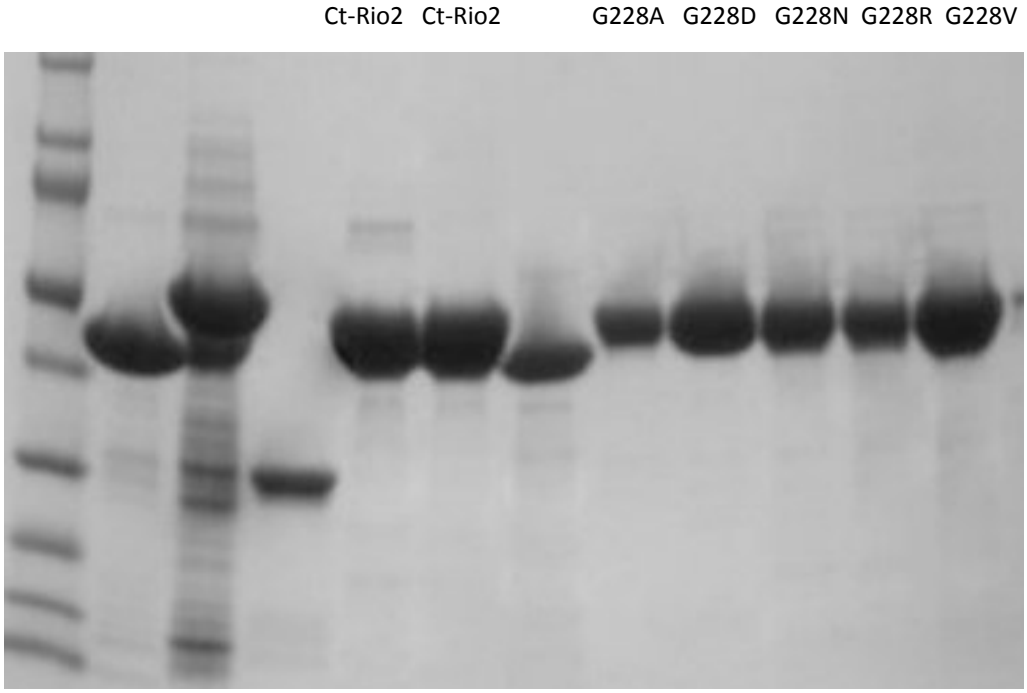

Figure 6.2 Preparative gels after size exclusion chromatography: Lane M is the protein ladder. Gel of Ct-Rio2 displays pure protein at the expected molecular weight of 44.6 kDa for all of the mutants.

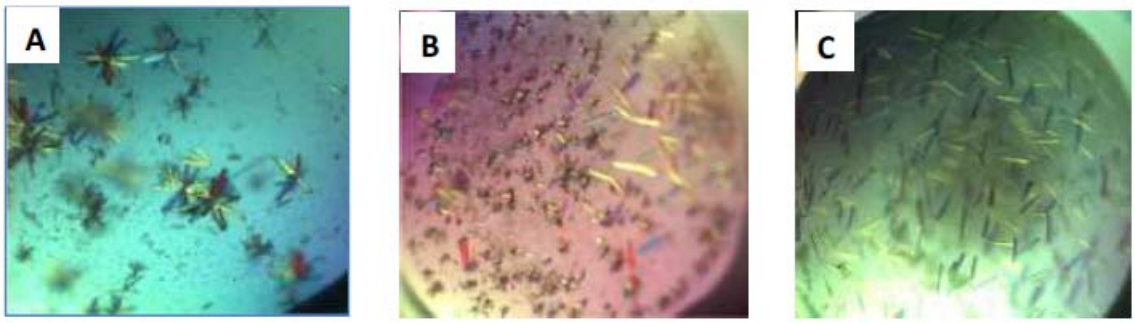

Figure 6.3 Ct-Rio2 mutants' crystals: A. G228A **B.** G228N and **C.** G228D Grown with ATP/Mg^{2+}.

Table 6.1 Data collection and current refinement statistics for AMP bound G228N. The Ct-Rio2 crystals formed in the presence of ATP in the hanging drop at pH=5.4.

Resolution range (Å)	59.27 - 2.26 (2.341 - 2.26)
Space group	R 3:H
Unit cell	118.623 118.623 72.556 90 90 120
Total reflections	91917 (9664)
Unique reflections	17775 (1835)
Multiplicity	5.2 (5.3)
Completeness (%)	99.71 (99.67)
Mean I/sigma(I)	17.59 (1.58)
Wilson B-factor	57.85
R-merge	0.05927 (1.069)
R-meas	0.066
CC1/2	0.999 (0.576)
CC*	1 (0.855)
R-work	0.2003 (0.2789)
R-free	0.2469 (0.3160)
Number of non-hydrogen atoms	2723
macromolecules	269
ligands	32
Protein residues	343
RMS(bonds)	0.015
RMS(angles)	1.71
Ramachandran favored (%)	91
Ramachandran outliers (%)	3
Clashscore	16.52
Average B-factor	82.20

Table 6.2 Data collection and current refinement statistics for AMP bound G228A: The Ct-Rio2 crystals formed in the presence of ATP in the hanging drop at pH=5.4.

Resolution range (Å)	59.83 - 2.67 (2.766 - 2.671)
Space group	R 3 :H
Unit cell	119.658 119.658 72.342 90 90 120
Total reflections	14453 (329)
Unique reflections	10772 (224)
Multiplicity	2.5 (1.5)
Completeness (%)	98.10 (97.83)
Mean I/sigma(I)	9.51 (0.97)
Wilson B-factor	69.38
R-merge	0.572 (0.5181)
R-meas	0.6674
CC1/2	0.986 (0.663)
CC*	0.996 (0.893)
R-work	0.1865 (0.3083)
R-free	0.2578 (0.4007)
Number of non-hydrogen atoms	2587
macromolecules	2555
ligands	32
Protein residues	343
RMS(bonds)	0.009
RMS(angles)	1.36
Ramachandran favored (%)	90
Ramachandran outliers (%)	3
Clashscore	10.47
Average B-factor	84.70

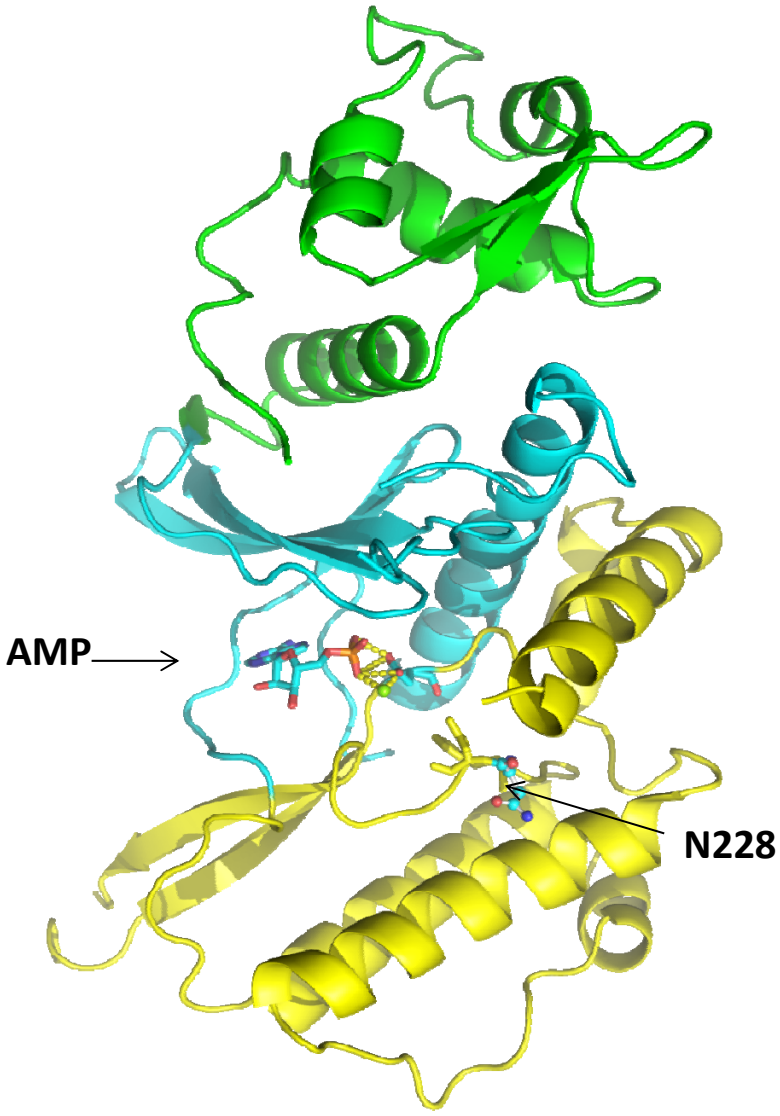

Figure 6.4 Crystal Structure of AMP bound G228N: The crystal structure of Ct-Rio2 with Asn mutation at residues 228 shows a bilobal kinase domain (blue and yellow) connected to an N-terminal winged helix domain (green). The "hinge" region connects the N- and C-terminal lobes of the kinase domain.

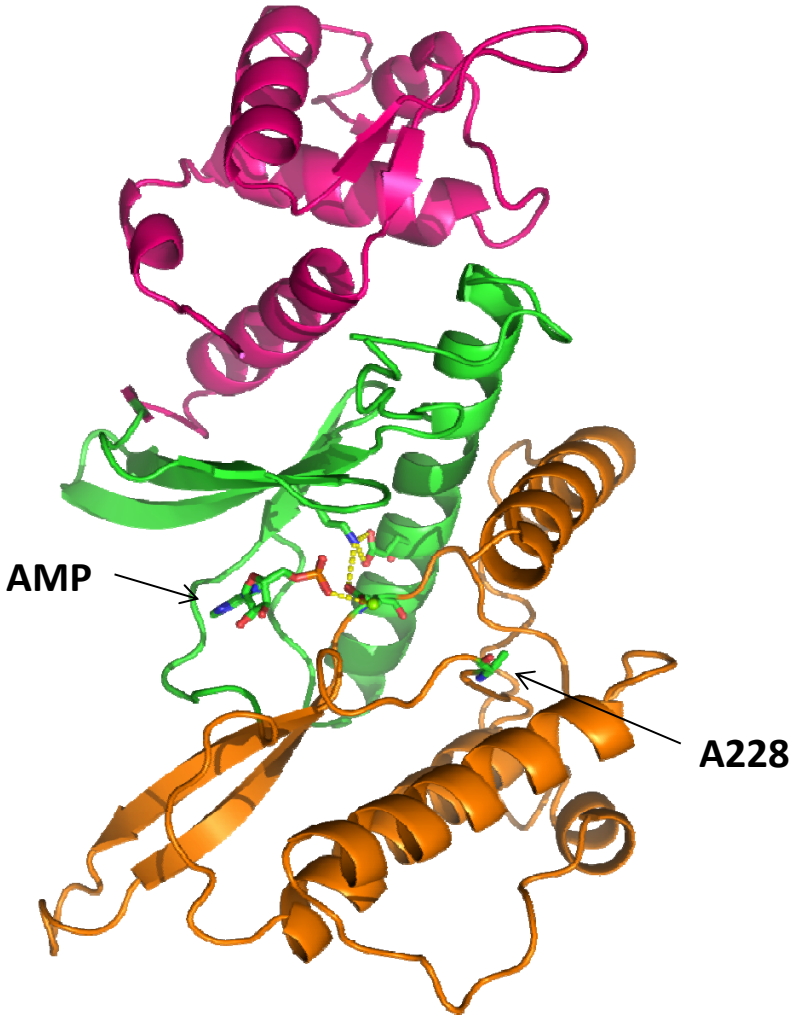

Figure 6.5 Crystal Structure of AMP bound G228A: The crystal structure of Ct-Rio2 with alanine mutation at residues 228 shows a bilobal kinase domain (green and orange) connected to an N-terminal winged helix domain (pink). The "hinge" region connects the N- and C-terminal lobes of the kinase domain.

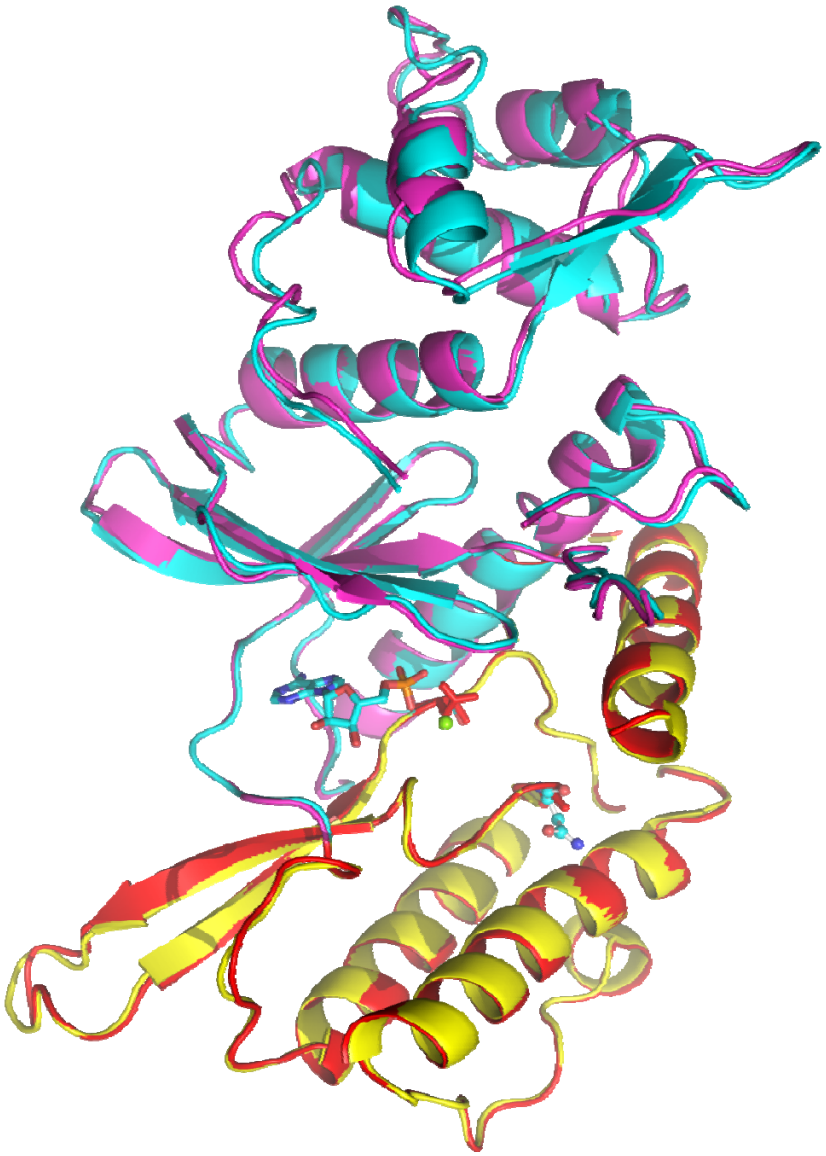

Figure 6.6 The overlay of the C-Terminal domain in the crystal structures of G228N and G228G (wild type Ct-Rio2): The winged helix domain of the two structures are not completely superimposable. The presence of Asn in the structure pushes the winged helix domain to provide a larger phosphate binding loop.

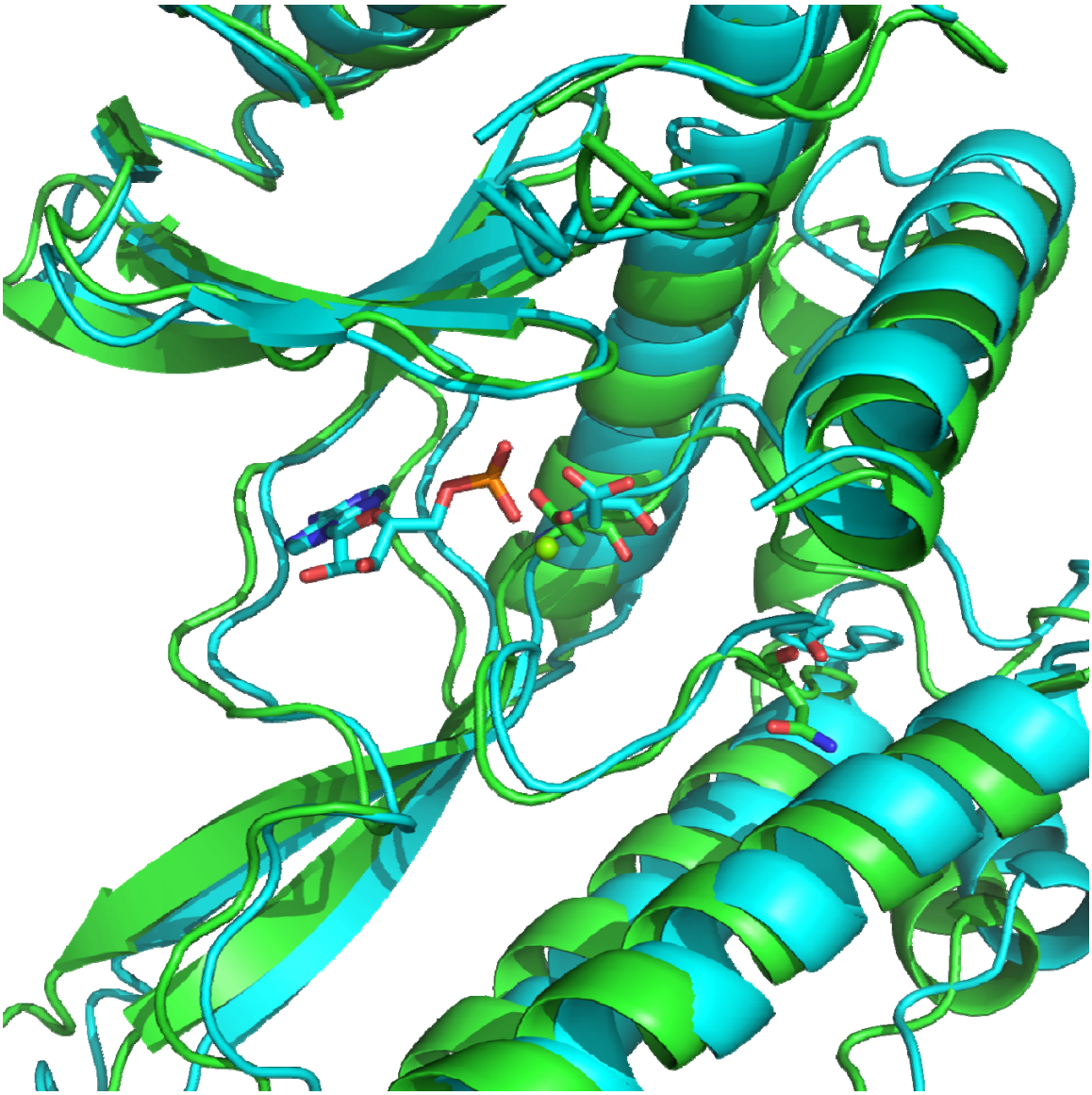

Figure 6.7 The overlay of the AMP domain in the crystal structures of G228N and G228G (wild type Ct-Rio2): The N and C terminal lobes of the two structures are not completely superimposable. The presence of Asn in the structure pushes the winged helix domain back to provide a larger phosphate binding loop.

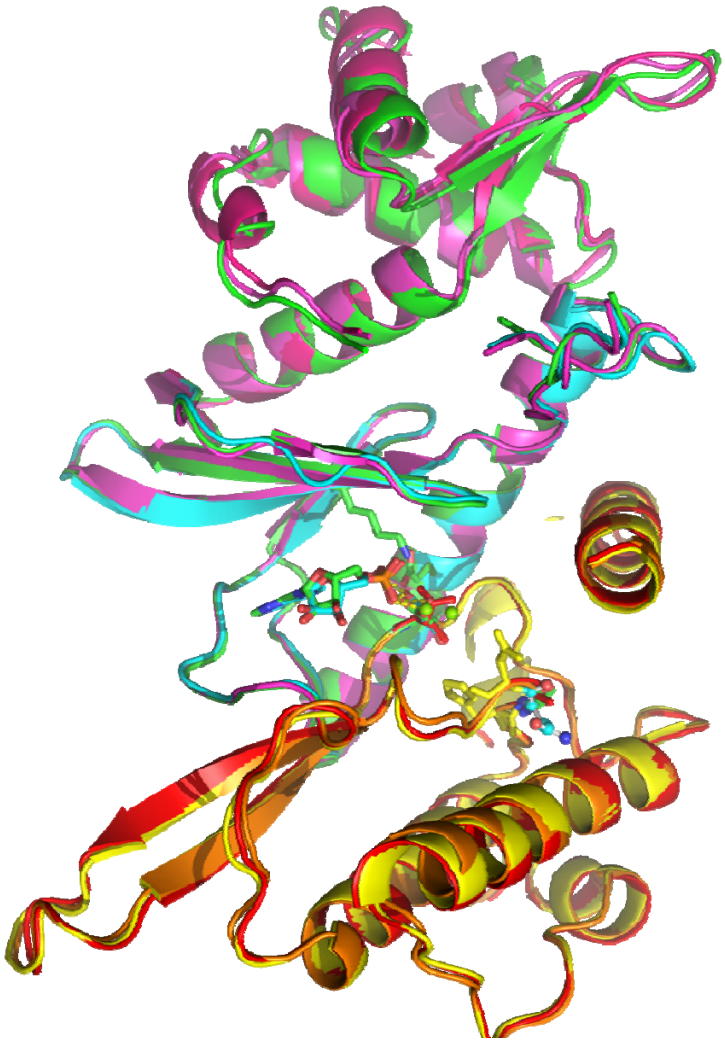

Figure 6.8 The overlay of the C-Terminal domain in the crystal structures of G228N, G228A and G228G (wild type Ct-Rio2): The winged helix domain of the G228A and G228 are very close to each other compared to the G228N. The presence of Asn in the structure seems to push the winged helix domain back to provide a larger phosphate binding loop.

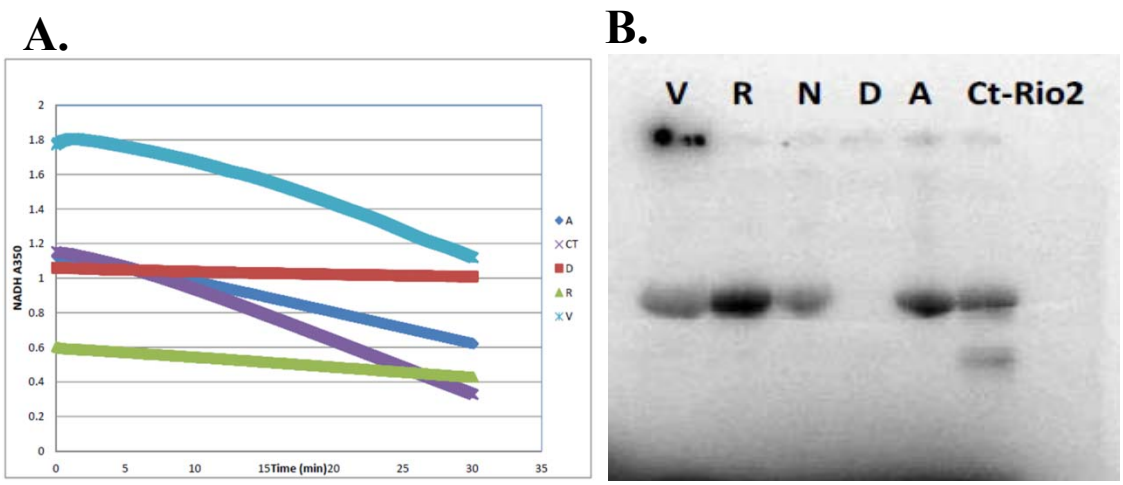

Figure 6.9 The proteins with different point mutations at 228, show very different activities in their A) ATPase and B) Auto-phosphorylation experiments: The G228A and Ct-Rio2 with G228 are similarly active in the hydrolysis of ATP but mutants residues such as those mostly observed in epks (G228V, and G228R) are less active. In line with our predictions, the auto-phsphorylation of Ct-Rio2 mutants follows an opposite order.

Chapter 7: Conclusion and Future Work

7.1 Crystal Structure of EspB

In chapter two of this dissertation, the purification and crystallization procedure of EspB is explained in detail. We optimized purification of EspB to obtain highly pure protein samples that were immediately used for crystallization. In chapter two, the X-ray crystallographic structure of the helical domains of EspB from M. Marinum was determined in a crystal form encompassing residues 1-292.

Going back to our objectives that are presented in chapter 1, now we can answer our questions about EspB.

- The molecular structure of EspB

The EspB structure is composed of 16 long, antiparallel coiled-coil helices in the asymmetric unit. These monomers are arranged into a set of four homotetramers. Two sets of these tetramers are arranged parallel to each other. Contact is observed between the C-terminal loop of one tetramer and the C-lobe helices of the adjacent tetramer.

- How do EspB monomers interact with one another?

Every homo-tetramer has two monomers (monomers A and B and monomers C and D) that are arranged anti-parallel to each other creating an interacting interface and a central helix bundle. Contact is observed between the C-lobe helices of two monomers within the tetramer. For instance monomer A interacts with monomer D and monomer B interacts with C via the C-lobe helices. The monomers have regions on the N- and C-terminus that lack electron density due to disorder and flexibility on these ends.

- How does EspB oligomeric state in the crystal relate to its biological functions?

The three dimensional structure of a protein provides insights into its probable function. Therefore, proteins with similar functions often tend to adopt similar folds even when they have very low sequence similarity. The results obtained from the x-ray crystallography of EspB indicate novel functional roles for the secreted proteins Rv3881c (EspB).

The 3D fold structure of EspB292 shows a helix-turn-helix fold at the N-terminal of the protein and the C terminal end of EspB292 has a WEG tripeptide motif at the position 176–178 that is similar to the WXG motif observed for the ESAT-6 like family of proteins. Based on previous results, the N-terminal region of Rv3879c displays helical regions followed by coiled-coil regions, similar to fibrinogen fold. These observations suggest that the N-terminal regions of EspB and Rv3879c can form a four-helix-bundle similar to CFP-10/ESAT-6 pair.

Interestingly, the 3D fold analysis of EspC protein also adopts a similar fold to that of ESAT-6 and EspB and the N-terminal region of EspA exhibited WXG motif. With regards to the secretory proteins of the ESX-1 secretion pathway, the current analysis indicates that the secretory protein pairs have a preference to adopt a four-helix-bundle structure. While the CFP-10/ESAT-6 and EspA/EspC pairs were previously reported to have such a structure, the present analysis suggests that the other secretory protein pairs EspB/Rv3879c may also adopt similar structures.

- How does the structure provide insight into cellular environments that Mtb encounters?

Based on our chemical cross-linking experimental observations and mass spectrometric studies, as well as EspB structure, EspB292 proteins form complex assemblies only at high concentrations (>100 μM). Therefore, the cleaved EspB protein that is secreted into the extracellular milieu is capabale of forming higher order oligomeric states and therefore, forming large protein assemblies when it accumulates in the small periplasmic space.

Since the truncation of C-terminus of EspB is necessary to regulate the secretion of virulence factors, the protein-protein tetramer network is probably competing with four-helix-bundles for translocation (through the mycomembrane) using the same machinery.

In mycobacterium tuberculosis' phagosomal lumen, iron plays a biphasic behavior. Its content first decreases up to four-fold within the first hours of infection, and increases from 5 to 25 folds during the following 24 hours. By seting up experiments that would mimic the phagosomal environment, we studied how the virulence factors are affected at each condition. To determine the effect of metal ions concentration on the EspB292 oligomerization, we compared the aggregation level of protein in various amounts of different metals. EspB exists mainly in its monomeric form in presence of many ions including Na^{+}, K^{+}, Cl^{-}, Ca^{2+}, Mg^{2+}, Po_4^{3-}, So_4^{2-} and Zn^{2+}. However, our findings further suggest that the aggregation of EspB292 can be significantly influenced by the content of the iron, and copper ions in the environment

- How does the structure of EspB predict the interaction with virulence factors?

Based on the crystal structure we can also observe that EspB can adopt a homodimer fold. The N terminal domain of EspB roughly resembles the helix-turn-helix domains of CFP10. The structure of EspB292, residues 135–226, is also roughly superimposable with the ESAT-6 structure with residues 2–104. This region of ESAT-6 and EspB each has a WxG motif at the turn of the second helix-turn-helix subdomain. In EspB292, these two subdomains form an extensive interface of non-covalent interactions, and the linker connecting them adopts an extended conformation that anchors loosely to the helical bundle through van der Waals interactions.

Notably, EspB also shows domains and regions that is specific to itself and is not found in ESAT-6 and CFP-10. For instance the helix 3, helix 6, and helix 7 of EspB292 create a

‘‘helical tip’’ that is rich in solvent-exposed hydrophobic/aromatic/acidic residues. The tip ends in an extended polyproline stretch (residues 270–288) that sandwiches between the subdomains through hydrophobic interactions provided by prolines and aliphatic side chains.

The EspB292 crystal structure also exhibit the unique relative disposition of multiple characteristic ESX sequence motifs, including a structured view of the helical export arm. In this work, we define the export arm to be residues 78–93 of EspB which is corresponding to the appendage that is dynamic in the structure. These residues from the export arm align on the same helical face upon binding the EccCb1 ATPase, targeting ESX secreted substrates for secretion.

- How does the structural information predict the EspB membrane transport mechanism?

Previous studies have provided experimental evidences that Rv3868 interacts with CFP-10/ESAT-6 and EspA/EspC pairs and helps in their multimerization. This multimeric protein assembly as well as EspB/Rv3879c is recognized by Rv3870 that is a member of Ftsk/SpoIIIE family of DNA translocase proteins. Rv3870 contains two transmembrane helices and a single FtsK ATPase domain and is known to interact with another cytoplasmic FtsK/SpoIIIE family protein Rv3871 that contains two FtsK ATPase domains. The α and β units of the FtsK domain containing proteins are known to multimerize to produce a hexameric ring with an inner diameter of 30Å. During cell division, this hexameric multimer acts as a DNA pump which translocates double stranded DNA with diameter of 20 Å. The α and β units domains of Rv3870/Rv3871 pairs also forms a hexameric ring, similar to that formed by a hexamer of FtsK proteins.

Structural analysis of CFP-10/ESAT-6 four-helix-bundle diameter has previously shown to be approximately 20–23 Å. Our findings suggest that each EspB292 monomer is similarly alpha helical and is roughly shaped like a cylinder with 103.6 Å length and a diameter of 16.1 Å that would perfectly fit and translocate through the hexameric ring of Rv3870/Rv3871 pairs. This

study suggest that, the FtsK family proteins (Rv3870 and Rv3871) recognize and translocate the four-helix substrates similar to that used by FtsK family proteins for the translocation of double stranded DNA. Therefore, the Rv3870/Rv3871 hexameric ring of T7SS may be a good example of machinery that has evolved from a specialized double stranded DNA translocator to a protein transporter.

The consecutive step after recognition of the secretory proteins by Rv3870/Rv3871 ring is translocation of these proteins through the inner membrane pore of Rv3877. The diameter of the inner pore considering a compact arrangement of the helices is determined to be approximately 26 Å or larger upon incorporating the side-chain lengths and interactions. Thus, this data also suggests that the substrate pairs (four-helix-bundles) of the ESX-1 secretion pathway would be able to fit into and be translocated along the inner pore of Rv3877.

7.2 Crystal Structure of Ct-Rio2 with AMP and ADP/Na_2VO_4

In chapters 4 and 5, the crystal structure of Ct-Rio2 was determined and the structures of AMP and ADP/VO_4 bound Ct-Rio2 were compared.

The following questions are answered based on the information that was obtained through crystallographic structure.

- What is the structure of Ct-Rio2 in presence of transition state analogs?
- What is the function of catalytic loop in the active site of this protein?
- What is the mechanism of the phosphoryl transfer from the ATP to the Asp residue?

In both of the structures of AMP- and ADP/VO_4-bound Ct-Rio2, the Ct-Rio2 protein showed a bilobal structure that was expected. In the AMP bound protein the Asp257 bonded the α-phosphate of the AMP in a different location and orientation compared to that of the ADP bound protein.

The movement of Asp257 in the transition state analog structure reveals that in the presence of the ATP the protein bilobal structure opens up to accommodate the ligand. The orientation of Asp257 is then changed to bind the gamma-phosphate of the ATP. In this case the vanadate structure mimics the γ-phosphate of the ATP. The structure of Ct-Rio2 in a complex with AMP aligns well with the previous apo protein structure. Moreover, Ct-Rio2 as expected showed an extended C-terminus containing αJ. The αJ interacts with the kinase domain through both hydrophobic and hydrophilic interactions with αC, αF, the flexible loop and the metal binding loop.

The Asp257 displays to both VO_4^{2-} and Mg^{2+} through polar interactions. In the bilobal kinase domain of Ct-Rio2 at pH 5.4, AMP is bound with no trace of VO_4 in its active site. The winged helix domain of the two structures is not completely superimposed. The overlay of the AMP region in the crystal structures of Ct-Rio2 bound to AMP and ADP at two different pH showed that both C and N terminal lobes were apart. The overlay of the active site of Ct-Rio2 bound to ADP and Orthovanadate vs. Ct-Rio2 bound to AMP displays a dramatic movement of Asp257 from the AMP bound complex to the vanadate bound complex.

7.3 The Evolutionary Origins of the Eukaryotic Protein Kinases

Another ultimate goal from the structural study of Ct-Rio2 was to quire knowledge about evolutionary relationship between Rio kinases and the canonical eukaryotic protein kinases. Towards this goal, the focus of my research was to target the non-ribosomal protein, Ct-Rio2 by asking the following questions:

- Can Rio2 act as the evolutionary origins of the eukaryotic protein kinases?

The procedure that was used for the purification of the mutants was at the pH of 5.4. We believe that the ATP that was originally mixed with the protein for crystallization has degraded

to AMP. But the comparison of the crystal structure of all of the mutants can be studied since all of them were attached to the same nucleotide. The overlay of the C-Terminal domain in the crystal structures of G228N and G228A of Ct-Rio2 exhibited that the winged regions were not superimposed perfectly and the presence of the Asn at the 228 opened up the phosphate binding loop. This could be the reason for the less ATPase activity of epks as compared to the Rio family of kinases.

The residues such as R, N and D are very similar in terms of charge and size to what we often observe in the eukaryotic protein kinases. But glycine and alanine that are found in Rio family of kinases close the catalytic loop and make the ATPase activity more preannounced. In contrast, the kinase activity seems to be less pronounced in the Rio family of the kinases. The epks perform kinase activity on the conserved serine residue much easier than Rio kinases. In fact, what we have observed so far suggest that Rio kinases are more of ATPases than kinases. Upon the overlay of the bound AMP groups none of the C and N terminal lobes neither the winged helix domain of the structures was completely superimposed. The overlay of the c-terminal region in the crystal structures of G228N, G228A and G228 of Ct-Rio2 bound to AMP showed that while the A and wild type Glycine Ct-Rio2 were superimposable the Asn mutated Ct-Rio2 was structurally different.

7.4 Future work

There are many interesting points yet to be answered about both of the EspB and Rio2 proteins. Our crystal structure of EspB along with further biochemical studies has revealed that EspB tends to aggregate to higher order oligomers. It would be interesting to find out how this oligomerization affects the virulence of Mtb in vivo. Our experiments also suggest that EspB similar to PE and PPE family of proteins has a helical tip that is perhaps performing an important

role in EspB binding to other proteins. This helical tip in PE.PPE class of proteins is known to bind a protein which is suggested to assist the proteins in secretion. We propose that EspB can be binding Rv3871 or Rv3879 genes. This is due to the experimental observation that shows in the absence if these to cytosolic proteins EspB are not capable of being transported to the extracellular milieu. This binding of these proteins to EspB can be studied both in vitro and in vivo followed by the structural study of EspB and Rv3871/79 proteins using crystallography. This piece of information can help us design therapeutics or drugs that would target the transport of EspB and hence affect the virulence by limiting the CFP-10 expressions as well as introducing defect in ESAT-6 and CFP-10 heterodimer secretion.

In terms of the Rio2 proteins there are many questions and problems that are yet to be answered. Our interest lies primarily in understanding the mechanism of phosphoryl transfer from ATP to the Asp257. Along with the structural studies that are listed in this thesis, the Rio protein's ATPase and phosphoryl release activity can be studied using kinetic analysis experiments. For example by enzyme coupled assays and stop flow experiment we can gain more specific insights about the mechanism of phosphate transfer and release to and from Rio2 protein.

References

1. Russell, D. G., *Nat Rev Microbiol* **2007,** *5* (1), 39-47.

2. Gouzy, A.; Nigou, J.; Gilleron, M.; Neyrolles, O.; Tailleux, L.; Gordon, S. V., *Res Microbiol* **2013,** *164* (3), 270-80.

3. Ernst, J. D.; Trevejo-Nunez, G.; Banaiee, N., *J Clin Invest* **2007,** *117* (7), 1738-45.

4. Zumla, A.; George, A.; Sharma, V.; Herbert, R. H.; Oxley, A.; Oliver, M.; Ilton, B. M. o., *Lancet Glob Health* **2015,** *3* (1), e10-2.

5. Serafini, A.; Boldrin, F.; Palù, G.; Manganelli, R., *J Bacteriol* **2009,** *191* (20), 6340-4.

6. Lewis, K. N.; Liao, R.; Guinn, K. M.; Hickey, M. J.; Smith, S.; Behr, M. A.; Sherman, D. R., *J Infect Dis* **2003,** *187* (1), 117-23.

7. Briken, V., *Curr Drug Targets* **2008,** *9* (2), 150-7.

8. Houben, D.; Demangel, C.; van Ingen, J.; Perez, J.; Baldeón, L.; Abdallah, A. M.; Caleechurn, L.; Bottai, D.; van Zon, M.; de Punder, K.; van der Laan, T.; Kant, A.; Bossers-de Vries, R.; Willemsen, P.; Bitter, W.; van Soolingen, D.; Brosch, R.; van der Wel, N.; Peters, P. J., *Cell Microbiol* **2012,** *14* (8), 1287-98.

9. Rathman, M.; Sjaastad, M. D.; Falkow, S., *Infect Immun* **1996,** *64* (7), 2765-73.

10. Hou, M.; Xie, J., *Wei Sheng Wu Xue Bao* **2010,** *50* (1), 15-22.

11. Das, C.; Ghosh, T. S.; Mande, S. S., *PLoS One* **2011,** *6* (11), e27980.

12. Welin, A.; Björnsdottir, H.; Winther, M.; Christenson, K.; Oprea, T.; Karlsson, A.; Forsman, H.; Dahlgren, C.; Bylund, J., *Infect Immun* **2015,** *83* (1), 205-13.

13. Francis, R. J.; Butler, R. E.; Stewart, G. R., *Cell Death Dis* **2014,** *5*, e1474.

14. Jo, E. K., *Pathog Dis* **2013,** *67* (2), 108-18.

15. Champion, P. A.; Stanley, S. A.; Champion, M. M.; Brown, E. J.; Cox, J. S., *Science* **2006,** *313* (5793), 1632-6.

16. Callahan, B.; Nguyen, K.; Collins, A.; Valdes, K.; Caplow, M.; Crossman, D. K.; Steyn, A. J.; Eisele, L.; Derbyshire, K. M., *J Bacteriol* **2010,** *192* (1), 326-35.

17. Ohol, Y. M.; Goetz, D. H.; Chan, K.; Shiloh, M. U.; Craik, C. S.; Cox, J. S., *Cell Host Microbe* **2010,** *7* (3), 210-20.

18. Renshaw, P. S.; Panagiotidou, P.; Whelan, A.; Gordon, S. V.; Hewinson, R. G.; Williamson, R. A.; Carr, M. D., *J Biol Chem* **2002,** *277* (24), 21598-603.

19. Wagner, J. M.; Evans, T. J.; Chen, J.; Zhu, H.; Houben, E. N.; Bitter, W.; Korotkov, K. V., *J Struct Biol* **2013,** *184* (2), 115-28.

20. Solomonson, M.; Huesgen, P. F.; Wasney, G. A.; Watanabe, N.; Gruninger, R. J.; Prehna, G.; Overall, C. M.; Strynadka, N. C., *J Biol Chem* **2013,** *288* (24), 17782-90.

21. Daleke, M. H.; Ummels, R.; Bawono, P.; Heringa, J.; Vandenbroucke-Grauls, C. M.; Luirink, J.; Bitter, W., *Proc Natl Acad Sci U S A* **2012,** *109* (28), 11342-7.

22. Brown, G. D.; Dave, J. A.; Gey van Pittius, N. C.; Stevens, L.; Ehlers, M. R.; Beyers, A. D., *Gene* **2000,** *254* (1-2), 147-55.

23. LAFONTAINE, D.; VANDENHAUTE, J.; TOLLERVEY, D., *Genes & Development* **1995,** *9* (20), 2470-2481.

24. Shajani, Z.; Sykes, M. T.; Williamson, J. R., *Annu Rev Biochem* **2011,** *80*, 501-26.

25. Nissan, T.; Bassler, J.; Petfalski, E.; Tollervey, D.; Hurt, E., *Embo Journal* **2002,** *21* (20), 5539-5547.

26. Sahasranaman, A.; Dembowski, J.; Strahler, J.; Andrews, P.; Maddock, J.; Woolford, J. L., *EMBO J* **2011,** *30* (19), 4020-32.

27. Granneman, S.; Baserga, S., *Experimental Cell Research* **2004,** *296* (1), 43-50.

28. Babiano, R.; Gamalinda, M.; Woolford, J. L.; de la Cruz, J., *Mol Cell Biol* **2012,** *32* (16), 3228-41.

29. Weis, B. L.; Missbach, S.; Marzi, J.; Bohnsack, M. T.; Schleiff, E., *Plant J* **2014,** *80* (6), 1043-56.

30. Venema, J.; BousquetAntonelli, C.; Gelugne, J.; CaizerguesFerrer, M.; Tollervey, D., *Molecular and Cellular Biology* **1997,** *17* (6), 3398-3407.

31. Ghalei, H.; Schaub, F. X.; Doherty, J. R.; Noguchi, Y.; Roush, W. R.; Cleveland, J. L.; Stroupe, M. E.; Karbstein, K., *J Cell Biol* **2015,** *208* (6), 745-59.

32. Wyler, E.; Zimmermann, M.; Widmann, B.; Gstaiger, M.; Pfannstiel, J.; Kutay, U.; Zemp, I., *RNA* **2011,** *17* (1), 189-200.

33. Grandi, P.; Rybin, V.; Bassler, J.; Petfalski, E.; Strauss, D.; Marzioch, M.; Schafer, T.; Kuster, B.; Tschochner, H.; Tollervey, D.; Gavin, A.; Hurt, E., *Molecular Cell* **2002,** *10* (1), 105-115.

34. Jakob, S.; Ohmayer, U.; Neueder, A.; Hierlmeier, T.; Perez-Fernandez, J.; Hochmuth, E.; Deutzmann, R.; Griesenbeck, J.; Tschochner, H.; Milkereit, P., *PLoS One* **2012,** *7* (3), e32552.

35. Sharma, K.; Venema, J.; Tollervey, D., *Rna-a Publication of the Rna Society* **1999,** *5* (5), 678-686.

36. Lebreton, A.; Rousselle, J. C.; Lenormand, P.; Namane, A.; Jacquier, A.; Fromont-Racine, M.; Saveanu, C., *Nucleic Acids Res* **2008,** *36* (15), 4988-99.

37. Zemp, I.; Wandrey, F.; Rao, S.; Ashiono, C.; Wyler, E.; Montellese, C.; Kutay, U., *J Cell Sci* **2014,** *127* (Pt 6), 1242-53.

38. Granneman, S.; Petfalski, E.; Swiatkowska, A.; Tollervey, D., *EMBO J* **2010,** *29* (12), 2026-36.

39. Campbell, M. G.; Karbstein, K., *PLoS One* **2011,** *6* (1), e16194.

40. Fassio, C. A.; Schofield, B. J.; Seiser, R. M.; Johnson, A. W.; Lycan, D. E., *Genetics* **2010,** *185* (1), 199-209.

41. Allmang, C.; Henry, Y.; Morrissey, J.; Wood, H.; Petfalski, E.; Tollervey, D., *Rna-a Publication of the Rna Society* **1996,** *2* (1), 63-73.

42. Bassler, J.; Paternoga, H.; Holdermann, I.; Thoms, M.; Granneman, S.; Barrio-Garcia, C.; Nyarko, A.; Stier, G.; Clark, S. A.; Schraivogel, D.; Kallas, M.; Beckmann, R.; Tollervey, D.; Barbar, E.; Sinning, I.; Hurt, E., *J Cell Biol* **2014,** *207* (4), 481-98.

43. Hector, R.; Burlacu, E.; Aitken, S.; Le Bihan, T.; Tuijtel, M.; Zaplatina, A.; Cook, A.; Granneman, S., *Nucleic Acids Research* **2014,** *42* (19), 12138-12154.

44. Schafer, T.; Maco, B.; Petfalski, E.; Tollervey, D.; Bottcher, B.; Aebi, U.; Hurt, E., *Nature* **2006,** *441* (7093), 651-655.

45. Jia, M. Z.; Horita, S.; Nagata, K.; Tanokura, M., *J Mol Biol* **2010,** *398* (5), 774-85.

46. Turowski, T. W.; Lebaron, S.; Zhang, E.; Peil, L.; Dudnakova, T.; Petfalski, E.; Granneman, S.; Rappsilber, J.; Tollervey, D., *Nucleic Acids Res* **2014,** *42* (19), 12189-99.

47. Chouhan, D. K.; Sharon, A.; Bal, C., *J Mol Model* **2013,** *19* (2), 485-96.

48. Ferreira-Cerca, S.; Sagar, V.; Schäfer, T.; Diop, M.; Wesseling, A. M.; Lu, H.; Chai, E.; Hurt, E.; LaRonde-LeBlanc, N., *Nat Struct Mol Biol* **2012,** *19* (12), 1316-23.

49. Ferreira-Cerca, S.; Kiburu, I.; Thomson, E.; LaRonde, N.; Hurt, E., *Nucleic Acids Res* **2014,** *42* (13), 8635-47.

50. LaRonde-LeBlanc, N.; Wlodawer, A., *J Biol Chem* **2005,** *280* (45), 37297-300.

51. Ray, W. J.; Puvathingal, J. M., *Biochemistry* **1990,** *29* (11), 2790-801.

52. CROWFOOT, D.; DUNITZ, J. D., *Nature* **1948,** *162* (4120), 608.

53. Fermi, G.; Perutz, M. F.; Shaanan, B.; Fourme, R., *J Mol Biol* **1984,** *175* (2), 159-74.

54. Gey Van Pittius, N. C.; Gamieldien, J.; Hide, W.; Brown, G. D.; Siezen, R. J.; Beyers, A. D., *Genome Biol* **2001,** *2* (10), Research0044.

55. Huang, D.; Bao, L., *J Microbiol Immunol Infect* **2014**.

56. Vandal, O. H.; Pierini, L. M.; Schnappinger, D.; Nathan, C. F.; Ehrt, S., *Nat Med* **2008,** *14* (8), 849-54.

57. Solomonson, M.; Setiaputra, D.; Makepeace, K. A.; Lameignere, E.; Petrotchenko, E. V.; Conrady, D. G.; Bergeron, J. R.; Vuckovic, M.; DiMaio, F.; Borchers, C. H.; Yip, C. K.; Strynadka, N. C., *Structure* **2015,** *23* (3), 571-83.

58. Dave, J. A.; Gey van Pittius, N. C.; Beyers, A. D.; Ehlers, M. R.; Brown, G. D., *BMC Microbiol* **2002,** *2*, 30.

59. Frasinyuk, M. S.; Kwiatkowski, S.; Wagner, J. M.; Evans, T. J.; Reed, R. W.; Korotkov, K. V.; Watt, D. S., *Bioorg Med Chem Lett* **2014,** *24* (15), 3546-8; Sun, D.; Liu, Q.; He, Y.; Wang, C.; Wu, F.; Tian, C.; Zang, J., *Protein Cell* **2013,** *4* (12), 921-31.

60. McLaughlin, B.; Chon, J. S.; MacGurn, J. A.; Carlsson, F.; Cheng, T. L.; Cox, J. S.; Brown, E. J., *PLoS Pathog* **2007,** *3* (8), e105; Chen, J. M.; Zhang, M.; Rybniker, J.; Boy-Rottger, S.; Dhar, N.; Pojer, F.; Cole, S. T., *Mol Microbiol* **2013**; Hamza, A.; Wagner, J. M.; Evans, T. J.; Frasinyuk, M. S.; Kwiatkowski, S.; Zhan, C. G.; Watt, D. S.; Korotkov, K. V., *J Chem Inf Model* **2014,** *54* (4), 1166-73.

61. Simon, S. R.; Konigsberg, W. H.; Bolton, W.; Perutz, M. F., *J Mol Biol* **1967,** *28* (3), 451-4.

62. De Leon, J.; Jiang, G.; Ma, Y.; Rubin, E.; Fortune, S.; Sun, J., *J Biol Chem* **2012,** *287* (53), 44184-91.

63. Atkinson, S. F.; Bettinger, T.; Seymour, L. W.; Behr, J. P.; Ward, C. M., *J Biol Chem* **2001,** *276* (30), 27930-5.

64. Soldati, T.; Neyrolles, O., *Traffic* **2012,** *13* (8), 1042-52.

65. Sturgill-Koszycki, S.; Schlesinger, P. H.; Chakraborty, P.; Haddix, P. L.; Collins, H. L.; Fok, A. K.; Allen, R. D.; Gluck, S. L.; Heuser, J.; Russell, D. G., *Science* **1994,** *263* (5147), 678-81.

66. Astarie-Dequeker, C.; Le Guyader, L.; Malaga, W.; Seaphanh, F. K.; Chalut, C.; Lopez, A.; Guilhot, C., *PLoS Pathog* **2009,** *5* (2), e1000289.

67. Hershko, C., *Baillieres Clin Haematol* **1994,** *7* (4), 965-1000.

68. Lounis, N.; Truffot-Pernot, C.; Grosset, J.; Gordeuk, V. R.; Boelaert, J. R., *J Clin Virol* **2001,** *20* (3), 123-6.

69. Wagner, D.; Maser, J.; Lai, B.; Cai, Z.; Barry, C. E., 3rd; Honer Zu Bentrup, K.; Russell, D. G.; Bermudez, L. E., *J Immunol* **2005,** *174* (3), 1491-500.

70. Serafini, A.; Pisu, D.; Palù, G.; Rodriguez, G. M.; Manganelli, R., *PLoS One* **2013,** *8* (10), e78351.

71. Cronje, L.; Edmondson, N.; Eisenach, K. D.; Bornman, L., *FEMS Immunol Med Microbiol* **2005,** *45* (2), 103-12; Meyer, D., *Curr Pharm Des* **2006,** *12* (16), 1943-7.

72. Russell, D. G., *Cell Host Microbe* **2011,** *10* (3), 181-3.

73. Hodgkinson, V.; Petris, M. J., *J Biol Chem* **2012,** *287* (17), 13549-55; Hodgkinson, V.; Petris, M. J., *J Biol Chem* **2012,** *287* (17), 13549-55.

74. Di, A.; Brown, M. E.; Deriy, L. V.; Li, C.; Szeto, F. L.; Chen, Y.; Huang, P.; Tong, J.; Naren, A. P.; Bindokas, V.; Palfrey, H. C.; Nelson, D. J., *Nat Cell Biol* **2006,** *8* (9), 933-44; Jentsch, T. J., Chloride and the endosomal-lysosomal pathway: emerging roles of CLC chloride transporters. In *J Physiol*, England, 2007; Vol. 578, pp 633-40.

75. Hershko, C., *Baillieres Clin Haematol* **1994,** *7* (4), 965-1000.

76. Widmann, B.; Wandrey, F.; Badertscher, L.; Wyler, E.; Pfannstiel, J.; Zemp, I.; Kutay, U., *Mol Biol Cell* **2012,** *23* (1), 22-35.

77. Hu, M.; Laronde-Leblanc, N.; Sternberg, P. W.; Gasser, R. B., *Parasit Vectors* **2008,** *1* (1), 34.

78. LaRonde, N. A., *J Biol Chem* **2014,** *289* (14), 9488-92.

79. LaRonde-LeBlanc, N.; Guszczynski, T.; Copeland, T.; Wlodawer, A., *FEBS J* **2005,** *272* (11), 2800-10.

80. Miranda-Saavedra, D.; Barton, G. J., *Proteins* **2007,** *68* (4), 893-914.

81. Geerlings, T. H.; Faber, A. W.; Bister, M. D.; Vos, J. C.; Raué, H. A., *J Biol Chem* **2003,** *278* (25), 22537-45.

82. Heo, J. B.; Lee, Y. M.; Yun, H. R.; Im, C. H.; Lee, Y. S.; Yi, Y. B.; Kwon, C.; Lim, J.; Bahk, J. D., *J Plant Physiol* **2014,** *171* (17), 1601-8.